THE UNIVERSITY *of* EDINBURGH

Identification of genetic influences in late-onset Alzheimer's disease (LOAD)

Mariet Allen BSc

Doctor of Philosophy
University of Edinburgh
2010

Declaration

I, Mariet Allen, hereby declare that the work presented in this thesis was performed by me. Exceptions to this are clearly stated in the text and include the collection of all samples and genotyping of all markers and SNPs used in the genome-wide linkage and association analyses. Measurement of plasma Aß for samples from the island of Vis was done co-operatively by Dr Morad Ansari (MRC Human genetic Unit, Edinburgh, UK) and I. Dr Linda Younkin (Mayo Clinic Jacksonville, FL, USA) also provided guidance and training for these experiments. The Quality Control script implemented within R prior to analysis of data with GenABEL was provided by Jennifer Huffman (MRC Human genetic Unit, Edinburgh, UK). Likewise, IBD estimates implemented in Loki, prior to linkage analysis in SOLAR, was provided by Dr Veronique Vitart (MRC Human genetic Unit, Edinburgh, UK).

The work presented in chapter 3 generally represents data from a publication for which I was an equal contributing author. Analysis of plasma Aß phenotypes was performed by me. Analysis of gene expression phenotype(s) was largely carried out by Dr Fanggeng Zou, Dr Minerva Carrasquillo and Dr Nilufer Ertekin-Taner (Mayo Clinic Jacksonville, FL, USA). Genotyping of AD and control subjects was mostly directed by Dr Minerva Carrasquillo, where I and others executed genotyping of multiple SNPs in a large number of subjects. Analysis of SNPs using a dual-luciferase assay was carried out by Dr Olivia Belbin (Mayo Clinic Jacksonville, FL, USA). Finally, sequencing experiments were carried out by me except where stated.

Acknowledgements

I would like to thank Professor Steven Younkin and Professor Alan Wright for being my supervisors. Their advice and guidance has been invaluable and will be something I will continue to return to in the future.

I would also like to thank my supervisors (above) and Dr Kate Wilson for agreeing to embark on this ambitious collaborative project with me.

I am grateful to Dr Linda Younkin for sharing with me your wisdom and expertise in measuring plasma Aβ and to Dr Morad Ansari for embarking on the plasma Aβ journey with me.

Thanks to Dr Veronique Vitart and Dr Caroline Hayward for your help with understanding the isolated populations evaluated in this thesis.

I would like to thank all of my friends and colleagues at the Mayo Clinic, Jacksonville and the MRC Human Genetics Unit in Edinburgh for providing me with technical support, friendship and laughs. Especially Chloe Stanton who I had the pleasure of living with during my time in Edinburgh and Minerva Carrasquillo who was always a listening ear.

I am especially grateful to the people from the islands of Vis and Korcula and the AD patients and their families who provided their samples for this study.

Finally I would like to thank my family who have always provided unconditional support throughout this process. A special thanks to Curt, Mum, Dad and Kate, I couldn't have done it without you!

Abstract

Late-onset Alzheimer's disease (LOAD) is the most common form of dementia, with an incidence of up to 50% in western populations over the age of 85 and a high heritability (up to 80%). The identification of risk factors for the development of LOAD is imperative for improving our understanding of this disease and for identifying therapeutic targets for treatment or prevention. Currently, the major known risk factors for the development of LOAD are age and the *ApoE* ε4 genotype. Previous studies have implicated plasma levels of the amyloid beta (Aß) peptide as a LOAD-associated quantitative trait and identification of loci influencing this trait could provide new insights into LOAD. In this thesis, plasma levels of the Aß peptides Aß40 and Aß42 have been measured in two isolated populations and genome-wide linkage and association analyses were performed. The genome-wide association analyses identified a number of promising quantitative trait loci; highlighting both novel and previously reported LOAD genes for further study, whilst also providing an excellent resource for genetic convergence studies with other LOAD related traits.

Several studies have reported an association between levels of oxidative stress and levels of Aß such that increasing levels of Aß appear to increase markers for oxidative stress and vice versa. The role of oxidative stress in LOAD and aging was therefore also investigated through analysis of mitochondrial mutational burden and DNA damage respectively, using DNA isolated from both blood and the brain and by carrying out a candidate gene association study of loci involved in mitochondrial function in LOAD cases and controls. Approaches to the investigation of mitochondrial genetics for the study of LOAD are comprehensively reviewed, adapted and tested and the results indicate a need for additional research in this aspect of the disease.

This thesis therefore presents a focus on two aspects of genetic research into LOAD, a complex disease with multiple environmental and genetic influences which aims to advance our understanding of the disease and bring us closer to treatment and prevention strategies.

Table of contents

List of figures

List of tables

List of abbreviations

ACh	Acetylcholine
AChE	Acetylcholinesterase
AD	Alzheimer's Disease
ADAM	A Disintegrin and Metalloprotease; refers to a family of proteases
ADRDA	Alzheimer's Disease and Related Disorders Association
AICD	APP intracellular domain generated by γ secretase cleavage
AlzGene	Website providing dabatase of published genetic association studies of AD and meta-analyses of these studies
APH-1	Anterior pharynx-defective 1, originally identied in C. Elegans; a subunit of the gamma secretase complex
APLP1	Gene encoding amyloid beta protein (Aß) precursor-like protein 1
APLP2	Gene encoding amyloid beta protein (Aß) precursor-like protein 2
APOE	Gene encoding Apolipoprotein E; "ApoE" refers to the protein
APOE ε4	*APOE* allele (2-SNP haplotype) associated with high risk of Alzheimer's disease
APP	Gene encoding amyloid beta protein (Aß) precursor protein; "APP" refers to the protein
APP-CTF	Carboxy-terminal fragment generated when APP is cleaved by α secretase (CTFα) or ß secretase (CTFß=CTF99).
APPswe	EOFAD-causing double (KM/NL) mutation of the two amino acids proximal to Aß1
Aß	Amyloid beta protein (peptide)
Aß1-40	Aß with amino acids 1-40.
Aß1-42	Aß with amino acids 1-42; Aß1-40 elongated by 2 amino acids at the carboxyl terminus.
Aß40	Aß peptide ending at the 40th amino acid in the Aß sequence, often used as shorthand for Aß1-40
Aß42	Aß peptide ending at the 42nd amino acid in the Aß sequence, often used as shorthand for Aß1-42
Aßx-40	The set of Aß species ending at amino acid 40 but with ragged amino termini e.g. Aß1-40, Aß3-40, etc.)
Aßx-42	The set of Aß species ending at amino acid 42 but with ragged amino termini (e.g. Aß1-42, Aß3-42, etc.)
BA-27	Aß40-specific detection antibody used in Aß sandwich ELISAs

BACE-1	Beta-site APP-cleaving enzyme 1
BACE-2	Beta-site APP-cleaving enzyme 2
BAN-50	Capture antibody used in Aß sandwich ELISAs, primarily captures Aß1-40 and Aß1-42
BC-05	Aß42-specific detection antibody used in Aß sandwich ELISAs
BNT-77	Capture antibody used in Aß sandwich ELISAs, captures Aßx-40 and Aßx-42
CAA	Cerebral amyloid angiopathy
CHRNB2	Gene encoding cholinergic receptor, nicotinic, beta 2
CO1	Gene encoding cytochrome oxidase 1
COII	Gene encoding cytochrome oxidase II
CSF	Cerebrospinal fluid
CTF99	The 99 amino acid carboxy-terminal fragment generated when ß secretase cleaves full-length APP; begins with Aß1-42 at its amino terminal.
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders IV
EOFAD	Early Onset Familial Alzheimer's Disease
eSNP	Single Nucleotide Polymorphism that associates with gene expression (transcript concentration in target tissue)
FASTA	Single letter text format for representing either nucleotide or protein sequences
FTD	Frontotemporal dementia
GenABEL	R package for large scale genetic analysis
GSAP	Gamma secretase activating protein
GWAS	Genome-wide association study
HVR	Highly variable region
IBD	Identity by descent
IDE	Gene encoding the insulin degrading enzyme
LD	Linkage disequilibrium
LOAD	Late Onset Alzheimer's Disease
LOC645503	Pseudogene similar to SUMO-1 activating enzyme subunit 2
LRP	Drug resistance gene, originally designated as encoding lung resistance-related protein

MAPT	Gene encoding microtubule associated protein tau; "tau" used to refer to the protein
MCI	Mild Cognitive Impairment
MCMC	Markov chain Monte Carlo, methodoglog for making IBD estimates
MMSE	Mini mental state examination
MRI	Magnetic resonance image/imaging
mtDNA	Mitochondrial DNA
NEDD9	Gene encoding "neural precursor cell expressed, developmentally down-regulated 9"
NFT	Neurofibrillary tangle
NIA	National Institute of Aging
NINCDS	National Institute of Neurological and Communicative Diseases and Stroke
NMDA	n-methyl D-aspartate, defines a specific type of glutatmate receptor
P3	Aß fragment generated from CTFα by γ secretase; equivalent to Aß17-40/42
PCR	Polmerase chain reaction
PEN2	Presenilin enhancer protein 1; a subunit of the gamma secretase complex
ProbABEL	R package for GWAS data imputation
PSEN1	Gene encoding presenilin 1
PSEN2	Gene encoding presenilin 2
PSP	Progressive supranuclear palsy
qPCR	Quantitiative PCR
RAGE	Receptor for advanced glycation end products
ROS	Reactive oxygen species
sAPPß	Large secreted N-terminal fragment generated by ß secretase cleavage
sAPPα	Large secreted N-terminal fragment generated by a secretase cleavage
SNP	Single Nucleotide Polymorphism
SOLAR	Sequential Oligogenic Linkage Analysis Routines, a software package for performing several kinds of statistical genetic analysis
SUMO-1	Small Ubiquitin-like Modifier 1
VaD	Vascular dementia

Chapter 1. Introduction

1.1 Overview

Alzheimer's disease (AD) was first described by Dr Alois Alzheimer in 1906 when he reported the clinical symptoms of a 51 year old female psychiatric patient (Auguste. D.) [4]. His post-mortem neuropathological examination of the patient's brain resulted in the first description of extracellular plaques and intracellular neurofibrillary tangles. Given the age of the patient, it is likely that she suffered from what is now referred to as early-onset Alzheimer's disease. Dr Alzheimer's findings in this first patient were presented at the 37th Meeting of the Southwest German Psychiatrists in Nov 1906 and published as a short report in 1907 [4, 5]. In 1910, Dr Alzheimer's colleague and mentor, Dr Emil Kraepelin, named the disease after him and a more extensive report was published in 1911, which also included the clinical and pathological description of a second patient, a 56 year old man (Johann.F.) [6, 7].

For many years it was believed that the disease described by Alzheimer was a rare form of pre-senile dementia. This changed in 1970 when Tomlinson, Blessed and Roth [8] showed that a substantial percentage of elderly patients with dementia have large numbers of plaques and tangles that are not distinct from those occurring in the presenile dementia described by Alois Alzheimer. Following this seminal observation, Alzheimer's disease has emerged as an important affliction of the elderly that is clearly separate from normal aging. It is relatively rare before the age of 65 years but, in industrialized countries, is present in ~10% of elderly subjects

over the age of 65 and up to 50% of people over the age of 85, where it is the most common form of dementia.

Alzheimer's disease is characterized by progressive memory loss, personality changes and eventual loss of motor functions, resulting in death. The range of life expectancy for patients following onset of symptoms is from 3 to 10 years. With the rapid growth of the elderly population, AD places an ever-increasing burden on health care systems and caregivers [9, 10]

1.2 Diagnosis of AD

Neurologists diagnose Alzheimer's disease based upon a clinical examination and detailed patient and family history in combination with patient performance on cognitive tests. Several tests have been developed for measuring the cognitive function of suspected AD patients. The most commonly used test is the mini-mental state examination (MMSE), although other more rigorous tests are also used. Specific criteria have been set forth by the NINCDS-ADRDA (National Institute of Neurological and Communicative Diseases and Stroke/Alzheimer's Disease and Related Disorders Association) [11] and the DSM-IV (Diagnostic and Statistical Manual of Mental Disorders) work group in order to ensure standardized diagnosis of Alzheimer's disease. Most neurologists adhere to these guidelines, which clinically separate patients into those with a diagnosis of "probable AD", where only evidence for AD is present, and "possible AD", where there is some evidence for other pathology that could cause dementia (e.g. cerebrovascular disease).

Although considerable progress has been made in accurately diagnosing AD during life, the only way to make a definite diagnosis in a demented patient is by demonstrating large numbers of senile plaques and neurofibrillary tangles in the brain at autopsy. Several papers recommending pathological guidelines for making a definite diagnosis of AD have been published, the most frequently used are the CERAD criteria for quantification of plaques, Braak staging for quantification of NFT's and the NIA-Reagan guidelines [12-14]. Despite the development of pathological guidelines for a definite diagnosis of AD, diagnosis is often complicated because pathological lesions associated with other dementias are present in addition to plaques and tangles. This not infrequently results in a post-mortem diagnosis of mixed dementia with AD [15].

Aside from clinical tests for the diagnosis of AD, Dr Ron Peterson at the Mayo Clinic developed criteria for determining incipient dementia termed Mild Cognitive Impairment (MCI) [16-18]. Within five years, approximately 50% of subjects diagnosed with MCI develop Alzheimer's disease. The diagnosis of MCI is, therefore, useful for identifying patients who will develop AD early in the course of disease, when therapy may be more effective.

In addition to the clinical methods that neurologists use for diagnosing MCI and probable or possible AD, researchers are continually working to develop more reliable and sophisticated methods for diagnosing AD at an early stage using neuroimaging and biochemical approaches. One good example is Pittsburgh compound B (PiB), a fluorescent analog of thioflavin T that binds to amyloid in senile plaques. PiB, can be used effectively in positron emission tomography scans

to image plaques during life [19-21] and is now being employed in many clinical studies of AD. Additional neuroimaging markers, biomarkers present in CSF or plasma and identification of genetic risk factors are all potential clinical tools that may be useful in the future for predicting risk of this disease.

1.3 Neuropathology

There are two classical neuropathological features that lead to a definite diagnosis of AD at autopsy. These are extracellular amyloid plaques (sometimes referred to as senile or neuritic plaques) and intracellular neurofibrillary tangles (NFT's) (Figure 1.1). In addition, there is substantial neuronal loss in AD that results in atrophy of the brain and widespread synaptic loss.

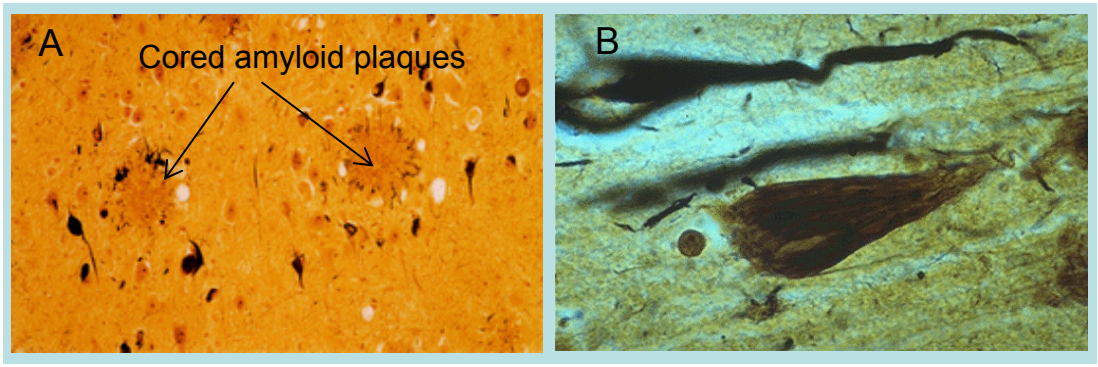

Figure 1.1 Neuritic plaques and neurofibrillary tangles in Alzheimer's disease brain. **A)** Modified Bielschowsky silver staining identifies dense neuritic plaques. Neuritic plaques are typically circular and 10-50 microns in diameter. **B)** High power view of two neurofibrillary tangles stained with a Bielschowsky silver stain. Images A and B provided by Dr Steven G. Younkin..

1.3.1 Neurofibrillary tangles and tau

In his initial report, Alzheimer used the Bielschowsky silver staining method to observe a then novel intracellular lesion. It is now known that the neurofibrillary tangles first described by Alzheimer are comprised of paired helical and some straight filaments in which the main component is the tau protein [22-26]. This pathology occurs in most cases of AD but it is not exclusive to AD. 'Tau neurofibrillary tangles' also occur in other neurodegenerative diseases collectively termed 'tauopathies', although the nature and distribution of these tangles varies depending on the disease.

The tau protein is a microtubule binding protein that stabilizes and promotes microtubule assembly. Tau is produced by the *MAPT* gene located on chromosome 17 and has six major isoforms expressed in the brain. In AD, tau aggregates within the cell, forming paired helical filaments when it is hyperphosphorylated, resulting in reduced microtubule stabilization, impaired axonal transport and intracellular tangles. [22, 27-29].

Neurofibrillary tangles can be identified using silver staining techniques similar to that used by Alzheimer, or more recently with immunohistochemistry techniques, using antibodies that recognize phosphorylated tau protein. The density and distribution of these tangles throughout the brain can be assigned to various categories called Braak stages, proposed and designated by Braak and Braak in 1991 [9]. This classification is used for post-mortem diagnosis of AD where a Braak stage of 4 or greater, combined with the presence of many dense amyloid plaques, indicates a diagnosis of AD.

In 1998, mutations at the *MAPT* locus were shown to cause one of the neurodegenerative diseases that fall under the classification of tauopathies, FTD17-P, so that it became clear that a loss of normal tau function (via mutation) can result in neurodegenerative disease [30]. In the case of AD there have, so far, not been any mutations identified at the *MAPT* locus that directly cause AD. However there have been at least 45 independent published AD case control association studies reported for SNPs at the *MAPT* locus with 15 of them reporting a positive association with AD (data from AlzGene [31] August 2010).

Whilst it is likely that a loss of normal function of the tau protein, via hyperphosphorylation, contributes to the pathogeneses of AD, it is currently not clear what causes this. It is likely that a number of complex pathways result in the hyperphosphorylation of tau in the manner associated with AD.

Under normal conditions, tau is a soluble protein which can be measured in samples of CSF. Researchers are exploring CSF tau measures as a clinical tool for disease risk prediction and in genetic association studies.

1.3.2 Amyloid Plaques

Senile plaques, which are most commonly associated with AD, are complex extracellular lesions. Each plaque consists of an amyloid core surrounded by altered neurites. The altered neurites, which contain paired helical filaments like those that comprise neurofibrillary tangles, can be identified with silver stains. Enveloping the cluster of altered neurites are astrocytes, with processes that project toward the amyloid core. Activated microglia are closely associated with most of the amyloid

core, indicating a possible role for inflammation in this disease [32]. This type of plaque, with a dense core, can be referred to as ‘dense plaques’. In most cases of AD, amyloid is deposited not only in these plaques but also in the walls of blood vessels in the brain, a disorder referred to as cerebral amyloid angiopathy (CAA).

Amyloid, by definition is highly insoluble extracellular protein in the form of 5-10 nm wide, relatively straight fibrils. The protein in amyloid fibrils has a crossed beta sheet conformation which binds aromatic dyes like Congo Red. Amyloid is deposited in various organs in many different diseases since there are many different proteins that can spontaneously assemble into insoluble amyloid fibrils in the extracellular space. In the mid 80s, the amyloid in AD brain and meningeal vessels was isolated, solubilised, and sequenced [33-35]. This amyloid was found to be composed of a then novel, ~4 kDa peptide, now referred to variously as the amyloid ß protein, amyloid ß peptide, amyloid ß, Aß, or A4.

When anti-Aß antibodies were used to immunostain AD brains, it became evident that, in addition to dense, cored plaques there was a second class of plaques referred to as diffuse plaques. The main constituent of diffuse plaques is also the amyloid ß protein, however these are frequently identified in both AD subjects and in the brains of elderly subjects without cognitive impairment. Diffuse Aß plaques can be found in areas of the brain not typically associated with AD such as the cerebellum and are not usually associated with abnormal neurites or reactive microglia and consequently may represent a pre-symptomatic form of Aβ deposition [36, 37].

Evidence from several studies suggests that relatively small Aß oligomers act as seeds which nucleate assembly of the large insoluble amyloid fibrils observed in

senile plaques. [38, 39]. Many proteins are associated with the fibrillar amyloid in senile plaques (Table 1.1). It is widely believed that these proteins adhere non-specifically and have little pathogenic significance, but could have implications for plaque formation and stability.

Surprisingly, the presence and density of senile plaques do not correlate well with severity of clinical symptoms [40-42]; this is in contrast with the other major AD pathology, NFTs. [12, 43].

Category	Examples
Proteoglycans	Heparan, chondroitin, keratin and dermatan sulphate proteoglycans.
Inflammatory molecules	Cytokines, chemokines, complement proteins, complement inhibitor, myeloperoxidase and immunoglobulins.
Metal Ions	Fe, Cu, Zn, Cr, Ni, Mn, Pb, Al and Si.
Amyloidogenic related molecules	Non-A component of AD amyloid (NAC), , low density lipoprotein receptor-related protein (LRP-1), cystatin-C.
Protease and clearance related molecules	1-antichymotrypsin, 1-trypsin, lysosomal proteinases, ubiquitin and α-2-macroglobulin.
Antioxidant defense proteins	Ferritin ceruloplasmin, Super OxideDismutase-1 & 2(SOD-1 & -2), heme oxygenenase (HO-1) and catalase.
Cholinesterase's	Acetylcholinesterase and butyrylcholinesterase.
Other proteins	Multifunctional clusterin (ApoJ, SP-40,40) apolipoprotein E (ApoE)

Table 1.1. Proteins and other molecules that have been reported to be associated with amyloid plaques in addition to the Aβ peptide. Adapted from Atwood.C.S *et al.* 2002 [44].

1.3.3 Atrophy of the brain

In AD, progressive neuronal and synaptic loss results in brain atrophy. This atrophy, which can be observed ante-mortem using *in vivo* imaging techniques such as magnetic resonance imaging (MRI) and upon post-mortem examination of the brain is also a characteristic feature of AD**.** Brain atrophy is observed in many forms of dementia. in each of which the type and location of cell death involved correlates with the clinical features of the disease [43, 45]. The temporal lobe, frontal lobe, and hippocampus are the areas most involved with cortical loss in Alzheimer's disease. Major association areas are also involved with notable cortical atrophy. Relatively spared areas include the primary motor and sensory cortex as well as the visual cortex. With atrophy of the brain parenchyma, there is notable enlargement of the cerebral ventricles. There is good evidence that neuroimaging techniques can identify atrophy of affected brain areas before the appearance of clinical symptoms indicating neuroimaging as a viable tool in the study and diagnosis of neurodegenerative diseases such as AD [46].

1.4 Amyloid-beta protein (Aβ)

In 1987, the gene that encodes Aβ was cloned and mapped to a locus on chromosome 21 [47-49]. In this gene (*APP*), Aß is encoded as an internal peptide within a much larger protein, the Aß precursor protein (APP). The APP is a transmembrane glycoprotein with a large luminal domain and a short cytoplasmic domain. The Aβ peptide sequence starts in the luminal domain and ends within the transmembrane domain. The APP gene produces three alternatively spliced

isoforms, but in each of them, Aß extends from just outside the membrane (amino end) to a position about half way through the single membrane-spanning domain (carboxyl end) [49, 50]. Interestingly, one of these isoforms (APP695) differs from the others in that it is largely neuron specific and lacks a Kunitz-type inhibitor domain that is common to the other two isoforms [51, 52]. The precise functions of these isoforms have not yet been elucidated. In mammals, there are two homologous proteins, called APP-like proteins (APLP1 and APLP2), although critically these lack the Aβ peptide domain.

When the *APP* gene was first cloned, it was known that patients with trisomy 21 (Down's syndrome) always develop AD pathology (senile plaques and neurofibrillary tangles) if they live past the age of 40 [53, 54]. This suggested that there is a gene, or perhaps a combination of genes, on chromosome 21, that can cause AD when over-expressed. When the *APP* gene was mapped to chromosome 21, it was evident that it might be the gene that causes AD in patients with trisomy 21.

In 1992, Dr. Younkin's laboratory and several others discovered essentially simultaneously that normal processing of the APP results in secretion of Aß [55-58]. Most secreted Aß has forty amino acids (Aß40) but a small percentage has two extra amino acids at its carboxyl end (Aß42) [59, 60]. Cells of all types produce APP and secrete both Aß40 and Aß42, which are readily detected in human plasma and cerebrospinal fluid [61, 62]. The cells in brain produce particularly large amounts, however, and the concentration of Aß in CSF is 50-times greater than in plasma .

This may explain why there is selective amyloid deposition and plaque formation in the brain as opposed to peripheral organs.

The normal physiological role, if any, of the amyloid ß protein is still currently not understood. Aβ is produced following enzymatic processing of the amyloid precursor protein (APP), known as amyloidogenic processing (see section 1.8, Figure 1.4). It initially exists as a monomer but under certain conditions Aβ can form oligomers that can aggregate into insoluble amyloid fibrils that are classically associated with AD [63, 64].

1.4.1 Neurotoxicity of Aβ

It is likely that Aβ exerts toxicity via multiple molecular mechanisms, an improved understanding of which can have important implications for the development of therapeutics for treating AD. The Aß42 peptide is particularly important in AD. In *vitro*, synthetic Aß42 spontaneously assembles into amyloid fibrils far more rapidly than Aß40 [63-65], so Aß42 is far more amyloidogenic than Aß40. In one third of AD brains, Aß ending at the 42 position is virtually the only form of Aß deposited and in another third it is by far the major form of Aß [66]. In the remaining third of AD brains, a large amount of Aß ending at position 40 is deposited in addition to Aß42. The amount of Aß42 deposited varies considerably from brain to brain, but appears to be independent of how much Aß40 is deposited. Thus, a large amount of Aß42 is deposited in all AD brains, whereas substantial amount of Aß40 is deposited in only one-third of AD cases [66]. Initially it was believed that aggregated amyloid deposits were the most neurotoxic species of Aβ. However more recent data

implicates both intra-cellular and extra-cellular Aβ oligomers as the most neurotoxic form of the peptide. In 1999, recently reviewed by Walsh D.M. and Selkoe D.J. [67], several research groups published data indicating improved correlation between soluble Aβ oligomers and disease course, implicating this as the neurotoxic species of aggregated Aβ [68-70]. Over the past decade there has been increasing evidence suggesting that soluble Aβ oligomers are indeed the pathogenic form of the peptide [67, 71, 72]. A thorough understanding of the neurotoxic effects of Aβ will have important implications for therapeutic targeting.

1.5 Genetics risk factors for Early Onset Familial AD (EOFAD)

Many families were identified in which AD occurred at an early age (35-60 years) with fully penetrant, autosomal dominant inheritance. Termed early onset familial AD (EOFAD), the AD produced through this simple Mendelian pattern of inheritance constituted a small fraction (<1%) of all AD, but analysis of the EOFAD families proved to be highly informative.

1.5.1 *APP*

Several families were identified in which EOFAD was produced by mutations in the *APP* gene. Many of the first mutations identified all occurred at the same location, namely, several amino acids from the carboxyl end of Aß [73-76]. The next *APP* mutations discovered included a double missense mutation (Swedish) affecting the two amino acids just before the amino terminus of Aß [77], as well as several additional mutations shown in Table 1.2. It is notable that several independent

mutations were found at the same residue of the APP protein, for example four different mutations cause the Val residue at position 717 in the APP protein to be altered. Furthermore, mutations in *APP* that cause AD tend to cluster in exons 16 and 17 of the transcript, which encode the Aβ peptide domain.

In addition to the mutations listed in Table 1.2, two studies also identified recessive mutations in the *APP* gene segregating in EOFAD families [78, 79]. Recently, duplication of the APP locus has been identified as causing EOAD with associated CAA [80, 81].

1.5.2 *PSEN1* and *PSEN2*

Following the identification of mutations at the *APP* locus on chromosome 21, researchers quickly realized that EOFAD was a heterogeneous disease since not all families could be explained by mutations in the *APP* gene. Mapping of additional loci that cause this disease was pursued and linkage to chromosome 14 was identified in some families [82-84]. In 1995, mutations at the *PSEN1* locus (Chr 14q24.3) were identified in EOFAD families with an aggressive form of the disease identified by a markedly earlier onset of clinical symptoms [85]. Over the past 15 years, many hundreds of variants have been identified at this locus in EOFAD pedigrees. These variants are distributed throughout the gene and do not appear to accumulate in any one domain of the protein. Due to the fact that so many variants have been described in this gene these have not documented here, but a comprehensive overview can be found at the AD & FTD mutation database: http://www.molgen.ua.ac.be/ADMutations [86].

Whilst mutations at the *PSEN1* locus were being investigated, a third EOFAD locus was also identified that was highly homologous to *PSEN1* [87, 88]. This is now known as the *PSEN2* gene and is located on chromosome 1q31-q42.

In contrast to *PSEN1*, variation at the *PSEN2* gene in EOFAD pedigrees is rarely observed, so that to date there are only 14 pathogenic mutations that have been found in this gene [86]. It is worth noting that the expression of *PSEN2* is lower than that of *PSEN1* in the brain and it is thought that Psen1 is the major presenilin that forms the γ-secretase in the brain [89-91]. This might explain why fewer mutations at the *PSEN2* locus cause EOFAD when compared with *PSEN1*.

Careful study of the genetic changes that cause AD, by the Younkin laboratory and others, showed that virtually all of these changes in *APP, PSEN1* and *PSEN2* increase Aß42. However, trisomy 21 increases both Aß42 and Aß40 and the double mutation in the *APP* gene described previously, which occurs in a large Swedish family, also increases both Aß42 and Aß40. These elevations in Aß40 and Aß42 have been demonstrated in cells transfected with mutant *PSEN1*, *PSEN2* and *APP* genes, the brains of transgenic mice expressing EOFAD mutations, and plasma from both symptomatic and presymptomatic subjects who have EOFAD mutations. [92-97]

Mutation	Alternate name	Exon/Domain	References	Year
Met 670 Leu, Lys 671 Asn	Swedish double	Exon 16, N-term	[77]	1992
Asp 678 Asn	Tottori	Exon 16, N-term	[98]	2004
Glu 682 Lys	Leuven	Exon 16, N-term	[99]	2008
Ala 692 Gly	Flemish	Exon 17, N-term	[100]	1992
Glu 693 Lys	Italian	Exon 17, N-term	[101]	1999
Glu 693 Gln	Dutch	Exon 17, N-term	[102-104]	1990
Glu 693 Gly	Arctic	Exon 17, N-term	[105]	1992
Asp 694 Asn	Iowa	Exon 17, N-term	[106]	2001
Ala 713 Thr	na	Exon 17, TM-1	[107-110]	2002
Thr 714 Ala	Iranian	Exon 17, TM-1	[111-113]	2002
Thr 714 Ile	Austrian	Exon 17, TM-1	[114, 115]	2000
Val 715 Met	French	Exon 17, TM-1	[116, 117]	1999
Val 715 Ala	German	Exon 17, TM-1	[118, 119]	2001
Ile 716 Phe	Florida	Exon 17, TM-1	[120]	1997
Ile 716 Phe	na	Exon 17, TM-1	[121]	2008
Ile 716 Thr	na	Exon 17, TM-1	[122]	2002
Val 717 Ile	London	Exon 17, TM-1	[74, 76, 123-128]	1991
Val 717 Phe	na	Exon 17, TM-1	[129, 130]	1991
Val 717 Leu	na	Exon 17, TM-1	[75]	2000
Val 717 Gly	na	Exon 17, TM-1	[73]	1991
Leu 723 Pro	Australian	Exon 17, TM-1	[131]	1998
Lys 724 Asn	Belgian	Exon 17, C-term	[132]	2006

Table 1.2 Pathogenic APP mutations segregating in families with EOFAD. Certain mutations are widely known by the geographical location of the primary family in which they were identified, noted here under 'alternate name'. The year of the first publication for each mutation is also indicated under 'year'. Adapted from: **http://www.molgen.ua.ac.be/ADMutations**, the AD & FTD mutation database [86].

1.6 Amyloid Cascade Hypothesis

One might argue that trisomy 21 and EOFAD mutations all foster Aß aggregation and plaque formation, but that these changes do not cause the neurofibrillary tangles, synapse loss, neuronal loss or dementia seen in AD. This line of thinking suggests that it is additional changes caused by these genetic variants, but not by elevated Aß42, that actually cause dementia. This seems highly unlikely, however, because an enormous body of evidence has accumulated indicating that aggregated Aß, most likely some form of soluble Aß oligomer, is highly toxic to neurons both *in vitro* and *in vivo* [67-70, 72]. In addition, there is good evidence that Aß accelerates the process of tangle formation in mice expressing the tau mutation (P301L) that causes fronto-temporal parkinsonism with dementia. This enhancement is particularly prominent in hippocampal pyramidal neurons which are widely believed to be most susceptible to tangle formation in AD [133].

Together, with the genetic evidence from the study of EOFAD, these observations indicate that increased production of Aβ can cause AD and it was from these findings that the amyloid cascade hypothesis for Alzheimer's disease was born [134-137]. The hypothesis postulates that the major triggering event in the disease is deposition of Aβ and that additional pathological and clinical changes occur as a result of Aβ deposition. This hypothesis is still held in high regard today, although it is widely believed that it is not insoluble amyloid fibrils but soluble Aβ oligomers that are the pathogenic form of Aβ in AD [67, 71].

As illustrated in Figure 1.2 there are two major pathways which can lead to an increase in Aβ in the brain: increased Aβ production or decreased Aβ clearance. This forms the basis of the amyloid cascade hypothesis.

As the evidence for the amyloid cascade hypothesis developed, many pharmaceutical and biotechnology companies began preclinical work to develop novel strategies to treat AD using Aß as the therapeutic target, and many academic laboratories began to study the trafficking and processing of APP and to investigate the mechanisms involved in removal of Aß. This effort generated an enormous amount of additional information that is summarized in the sections that follow.

1.7 APP trafficking

The APP protein has been identified at the plasma membrane implicating trafficking through the secretory pathway, but it has also been identified at the membranes of endosomes, lysosomes and mitochondria [138-140]. APP that reaches the plasma membrane through the secretory pathway is quickly internalized by recognition of a "YENPTY" internalization motif in APP. Following endocytosis of APP, much of it is believed to be delivered to lysosomes for degradation but a small component is recycled to the plasma membrane, as illustrated in Figure 1.3. This is believed to be a major pathway for Aβ secretion [141]. The cellular localization of APP has implications for proteolytic processing of the protein, which in turn has implications for the production of Aβ, consequently the cellular trafficking of APP is an area of intense research.

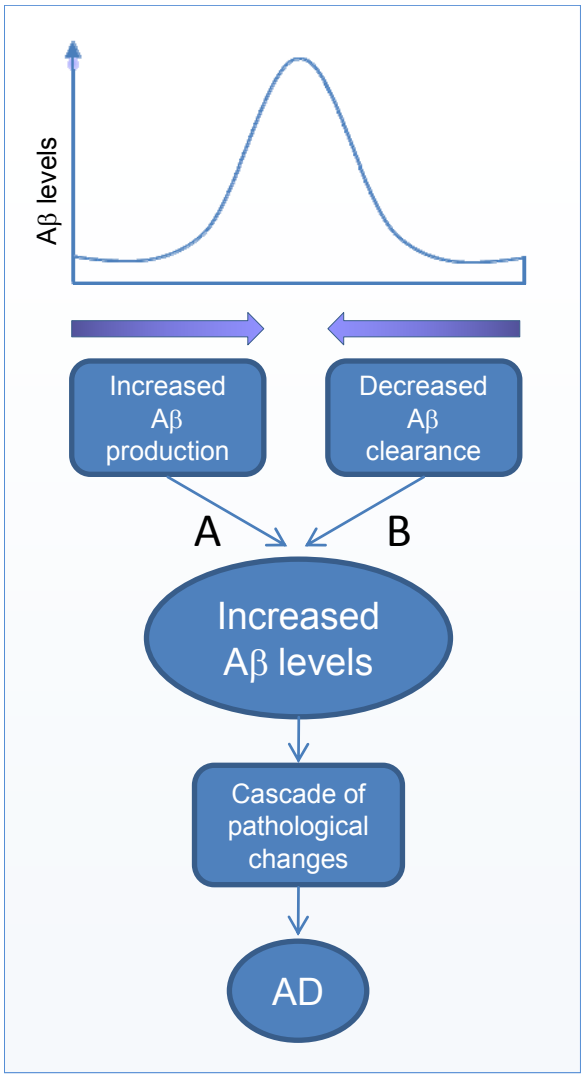

Figure 1.2. Interpretation of the amyloid cascade hypothesis.
This figure is a simplified interpretation of the amyloid cascade hypothesis, as multiple pathways may influence the cellular processes that in turn influence Aβ production and degradation. Studies of EOFAD however show that a direct effect of altered APP processing fits this hypothesis and it is therefore likely that the sporadic form of the disease also fits albeit a result of less direct processes.
A: Known mutations in APP, PS1 and PS2 that cause altered APP processing which result in either an increase in the total amount of Aβ produced or a relative increase in the more amyloidogenic form Aβ42.
B: Amyloid-beta degradation is less well defined, although multiple lines of evidence implicate several degradation pathways as being involved in AD pathogenesis. A number of amyloid-beta degrading enzymes have been identified.

For example, synaptic transmission is believed to lead to an increase in endocytosis of APP and consequently increased amyloidogenic processing (which is more prevalent in endosomes), resulting in increased levels of secreted Aβ [142-144]. Furthermore, the processing of APP appears to be sensitive to cellular cholesterol concentrations, since evidence from several studies suggests that an increased cholesterol content in membrane microdomains referred to as 'lipid rafts' fosters co-localization of APP and the secretases that release the Aβ peptide domain, resulting in an increased Aβ production [144, 145].

Multiple molecular mechanisms can influence the trafficking and proteolytic processing of APP, so a continued effort to understand these mechanisms may lead to improved therapeutic strategies.

1.8 APP processing

Three enzymes or enzyme complexes are known to cleave the APP protein, which are known as α, β and γ secretases. APP processing can follow an amyloidogenic pathway: cleavage by ß-secretase followed by γ-secretase. Or it can follow a non-amyloidogenic pathway: cleavage by α-secretase followed by γ-secretase. The amyloidogenic pathway results in the production of the Aß peptide (Figure 1.4) plus soluble APP-beta (sAPP-β) and the APP intracellular domain (AICD). The non-amyloidogenic pathway results in the production of soluble APP-alpha (sAPP-α), a 3-kDa Aβ derivative (P3) and AICD.

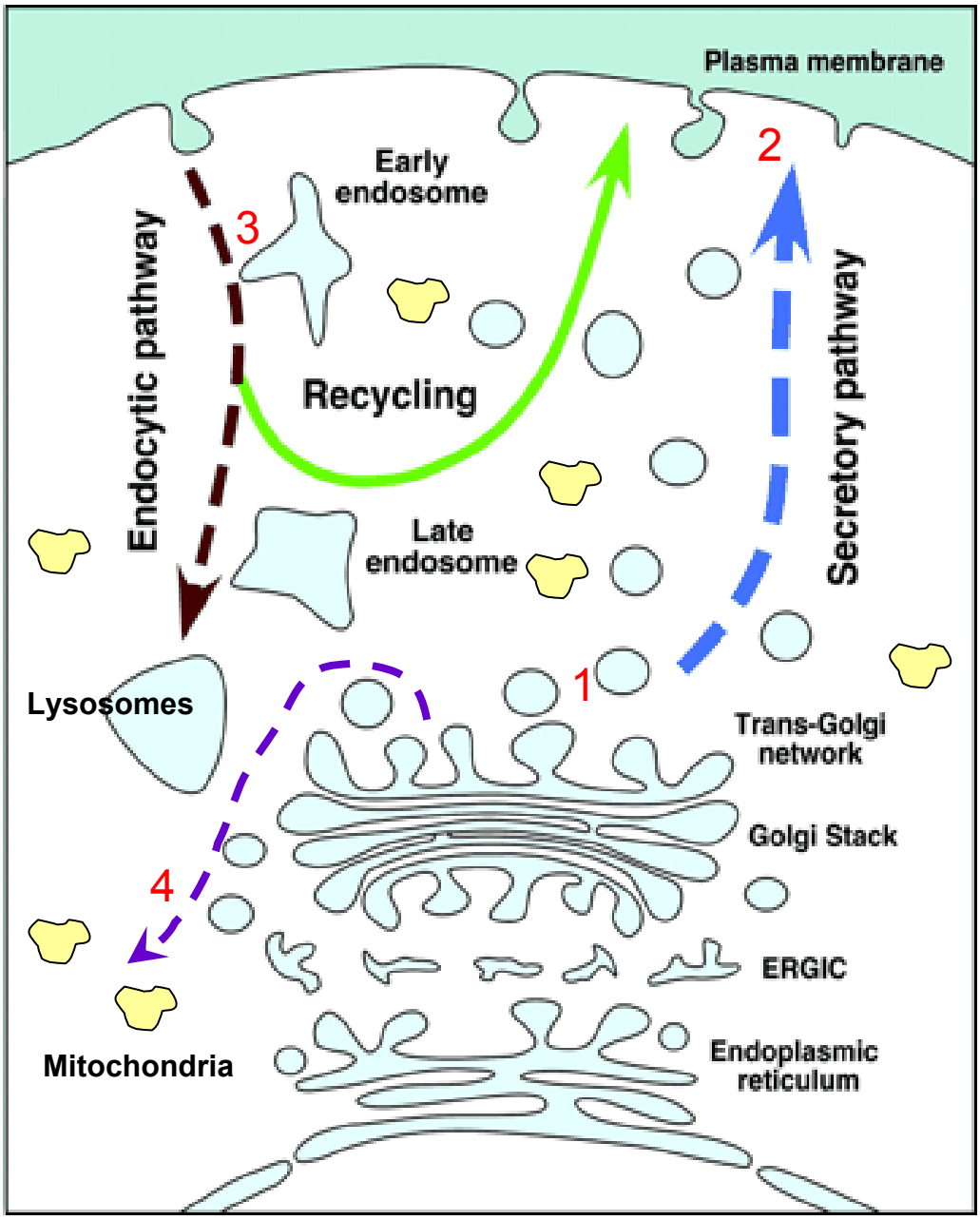

Figure 1.3. APP trafficking. 1. Anticipated site of both alpha and beta secretase activity. 2. Anticipated site of gamma secretase activity. 3. Expected gamma and beta secretase activity. 4. Evidence suggests that at least some APP is directed to mitochondrial membranes and may be cleaved here or continue on a secretary or lysosomal pathway. Figure adapted from Vetrivel and Thinakaran [138].

Similar to the Aβ peptide, P3 can vary in the C-terminal residues at the site of the gamma secretase cleavage, although the impact this may have on this peptide is poorly understood. Likewise, AICD will also have variability at its N-terminus due to heterogeneity at the γ-secretase cleavage site, however subsequent processing at a second site (ε-site) results in a 50 residue peptide, in which the variable length fragment is removed. Cleavage of APP, by either α or β secretase, results in the production of one of two forms of a long secreted APP N-terminal fragment (sAPPα and β), which differ at the carboxyl-terminal, with the alpha form containing the first 16 residues of the Aβ peptide. The putative roles identified for the products of APP processing, other than Aβ, are summarized in Table 1.3.

Studies of these proteins have identified various possible neuroprotective and neurotoxic effects, indicating that these may act as modifiers of Alzheimer's disease etiology. However, the presence of the Aβ peptide as the major constituent of the classic AD lesion, its propensity to aggregate and the wealth of evidence supporting its neurotoxic effects all indicate that Aβ is the main pathogenic peptide produced from the processing of APP in Alzheimer's disease.

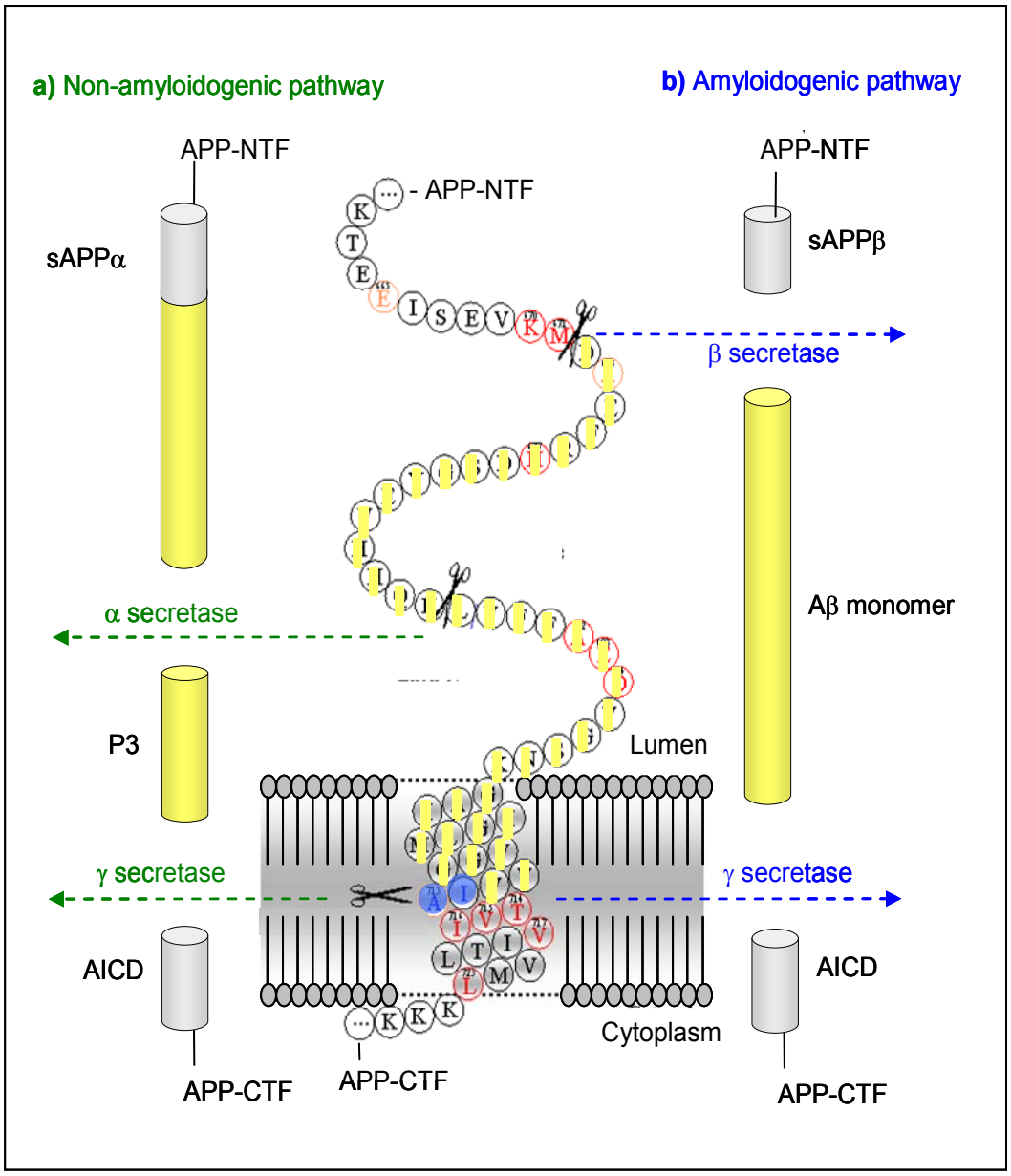

Figure 1.4 APP Processing: The APP is a transmembrane protein relevant to Alzheimer's disease, since proteolytic processing of APP results in the production of Aß protein, the major constituent of amyloid plaques. APP can be processed by ß-secretase followed by γ-secretase generating Aß (amyloidogenic pathway). Alternatively, it can be processed in a non-amyloidogenic pathway, with cleavage by α-secretase followed by γ-secretase. Mutations in the *APP* and *PSEN1/2* genes, which encode the catalytic subunit of γ-secretase, are associated with the early-onset familial Azheimer's disease. The Figure is modified from Hardy J. with blue circles representing residues 41 and 42. Red circles represent pathogenic *APP* mutations and orange circles representing APP mutations that are not likely to be pathogenic [146].

Protein/ peptide	Enzymes	Function related to the brain
sAPPα	alpha	Studies using rats indicate that sAPPα increases long term potentiation (LTP), and improves spatial memory [145], a finding that is inversely linked to some of the deficits observed in AD subjects. Others have shown positive effects, *in vitro,* of sAPPα on NMDA receptor associated neurite growth, synaptic plasticity and LTP [147-149], indicating that sAPPα may have neuroprotective properties.
sAPPβ	beta	Both secreted forms of the APP ectodomain can trigger an inflammatory response in the brain through activation of microglia via a glutamate pathway that might also interact with ApoE [150, 151]. However, unlike sAPPα, studies of sAPPβ have not identified any neuroprotective properties for this protein resulting in an imbalance in neurotoxic effects of sAPPβ that could consequently contribute to neuronal dysfunction and disease. Furthermore, *Li et al.* recently implicated a role for sAPPβ in transcriptional regulation of transthyretin and klotho proteins, which have been associated with the formation of β-amyloid aggregates and aging respectively [152-154].
P3	alpha & gamma	P3 is equivalent to Aβ 17-40/42 [155]. It has not been as extensively studied as full-length Aβ and its normal function is not well understood. P3 has been identified as the major constituent of extracellular deposits observed in the cerebellum of Down's Syndrome patients, called pre-amyloid lesions and it has also been found in AD neuritic plaques [156]. Studies have also implicated a role for P3 in neuronal apoptosis [157]. P3 probably represents a peptide with similar properties to Aβ, that has either reduced neurotoxicty or toxicity that is not well understood.
AICD	gamma	The APP Intracellular domain (AICD) is generated by both APP processing pathways (Figure 1.4.). Following initial γ secretase activity, the AICD is further cleaved at an ε cleavage site which results in the 50 residue AICD peptide. The AICD has been frequently linked to regulation of gene expression via association with various other factors such as Fe65 and Tip60. GSK3β, which has frequently been linked to tau hyper-phosphorylation, is among the genes believed to be regulated by the AICD [158]. Goodger Z.V. *et al.* have reported that the AICD generated via the amyloidogenic pathway has a greater preference for nuclear localization, and consequently regulation of gene expression, than that produced by the non-amyloidogenic pathway, perhaps due to the sub-cellular localization of the amyloidogenic pathway events [159].

Table 1.3. Summary of some of the putative roles identified for products of APP processing, other than Aβ.

1.8.1 Alpha secretase

As shown in Figure 1.4. alpha secretase cleaves APP between residues 16 and 17 of the Aβ peptide, resulting in the shedding of the APP ectodomain (sAPPα) and consequently precludes the production of full length Aβ. Recent studies indicate that members of the ADAM (a disintegrin and metalloprotease) family of proteases are responsible for α-secretase activity. This family of multi-domain, membrane bound proteases have been associated with the action of ectodomain shedding for a number of transmembrane proteins in addition to APP, although exactly which members of this family have this activity on APP in the brain remains to be elucidated. Several reports have identified the action of ADAM 9, 10, 17 and 19 as the most likely APP α-secretase candidates [160-163] whilst ADAM 9, 10 and 12 have been the target of AD genetic association studies, with varying results [164-168]. Furthermore the activity of ADAM proteases on APP has been identified at both the plasma membrane and the trans-golgi network [169]. This indicates multiple sub-cellular locations for α-secretase activity, which are consequently likely to be sensitive to multiple regulatory pathways. Collectively, the evidence to date suggests that whilst the APP alpha-secretase may be one or more members of the ADAM family, more research is needed to better understand how these functions might influence Alzheimer's disease.

1.8.2 Beta secretase.

Beta-secretase cleaves the APP protein resulting in the production of a 99 residue C-terminal fragment (CTF99) which is later cleaved by γ-secretase to produce

Aβ (Figure 1.4). In 1999, four research groups identified BACE-1, a membrane bound aspartyl protease, as having beta-secretase activity [85-88]. More recently, the same was shown for BACE-2 [89] and members of the cathepsin family of proteases [72, 90] although it appears that these proteases do not have the same level of beta secretase activity in the brain as BACE-1 and so probably do not have the same impact on the amyloidogenic processing of APP [72, 91]. APP is not the only substrate of BACE-1, since additional substrates include the APP like proteins (APLP-1 and 2), LRP and neuregulin [170]. Similar to the α and γ secretases, it is probable that BACE-1 acts upon these other substrates in sub-cellular compartments which may not entirely overlap with that of APP. BACE-1 performs optimally at low pH and is consequently most frequently found in acidic sub-cellular compartments such as endosomes [171], although evidence suggests that there is some activity at the plasma membrane and also in the Golgi. Cleavage of APP by β-secretase in intracellular compartments such as the Golgi is enhanced with the Swedish APP double mutation (APPswe), which is known to cause EOFAD (Table 1.2) [93, 172].

Studies of SNPs at the *BACE-1* locus for association with LOAD have revealed heterogeneous results [31], which is a common finding for LOAD candidate genes . Additional evidence implicating BACE-1 as the β-secretase comes from studies of BACE-1 knock-out mice, which exhibit an essentially complete depletion of neuronal Aβ secretion, providing further evidence that BACE-1 is critical for amyloidogenic APP processing [173].

The overwhelming majority of research indicates that BACE-1 is the most likely source of β-secretase activity. Additional enzymes may be able to cleave APP at the β-secretase cleavage site but it is unlikely that they are active *in vivo*.

1.8.3 Gamma secretase.

The gamma-secretase that cleaves APP is an aspartyl protease and, as illustrated in Figure 1.5, is a complex of four transmembrane proteins; Nicastrin, (NCT), Presenilin enhancer 2 (PEN2), anterior pharynx defective 1 (APH-1) and Presenilin 1 or 2 [89]. The presenilins are believed to be responsible for forming the catalytic domain and, in the brain, PS1 has been found to be the predominant presenilin [90, 91]. The remaining three cofactor proteins are thought to be involved in formation and stability of the complex, although some reports also indicate a role in substrate recognition [174, 175]. Current evidence indicates that γ-secretase cleaves APP within the membrane, where two aspartate residues, located within transmembrane domains 6 and 7 of presenilin, are critical to the hydrolysis of the target peptide bond [90]. This action is heterogeneous, since the particular bond that is hydrolyzed can vary, resulting in different terminal residues of the resultant peptides. Mutations in *APP*, *PSEN1* and *PSEN2* that have been implicated in EOFAD cause preferential hydrolysis at particular sites resulting in a selective increase in Aβ42 production. [73-75, 85, 87, 88].

The mature γ-secretase complex has been found in many sub-cellular compartments and is also found at the plasma membrane [140, 176, 177]. However, with respect to APP, the γ-secretase preferentially exerts its activity on the APP C terminal

fragment: APP-CTF (following cleavage by alpha or beta secretase) and is consequently most active against APP in compartments enriched in the APP-CTF.

The components of the γ-secretase are also known to have enzymatic activity against other type-1 membrane proteins, in addition to APP. The most significant and well known alternative γ-secretase substrate is Notch [178, 179]. For this reason, attempts at using the γ-secretase as a therapeutic target have so far failed. Notch is a signalling peptide involved in several cellular processes and consequently altered Notch processing can result in an embryonic lethal phenotype in mouse studies and unfavourable side-effects in drug trials [180-182]. A recent study identified loss-of-function mutations in *PSEN1* that cause a dominant form of the skin disease acne inversa (AI). It is likely that these mutations result in altered notch processing, the major target of the γ-secretase in skin cells [183]. Interestingly, Wang et al [183] indicate that AD and AI may be mutually exclusive suggesting that mutations of *PSEN1* and 2 that cause AD likely represent gain-of-function mutations affecting APP cleavage, thus altering the amount of Aß produced.

Recent developments have identified selective γ-secretase mechanisms that influence APP processing without affecting Notch. These include selective targeting of Aph-1 isoforms and more recently a novel γ-secretase activating protein (GSAP) [184, 185]. This provides promise for future research on the specific therapeutic targeting of γ-secretase effects on APP.

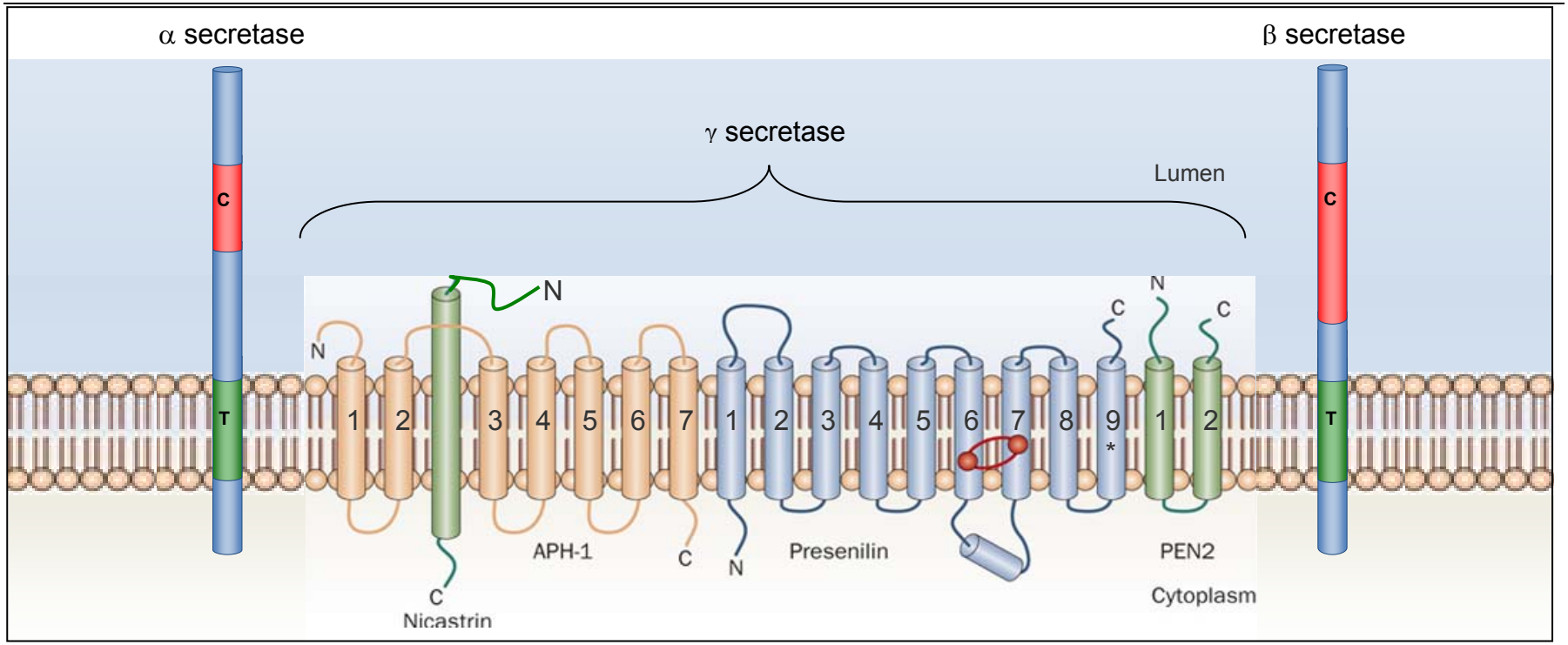

Figure 1.5. Postulated membrane location of α secretase β secretases, and the multi-subunit γ secretase. α and β secretases are believed to be membrane bound via a transmembrane domain (T) and to have a catalytic domain (C) as shown by the red domains. The catalytic domain of γ-secretase is believed to include two aspartate residues in transmembrane domains 6 and 7 of presenililin, as shown by the red circles. Two common models for presenilin are of 8 or 9 transmembrane domain proteins, although these models are not easily resolved, so the ninth transmembrane domain is shown here with an asterix. Modifed from De Strooper B. *et al.* [170].

1.9 Aβ degradation and clearance.

Understanding the molecular mechanisms involved in the production of the Aβ peptide is of critical importance in understanding the pathogenesis of AD. This is highlighted by the fact that most cases of EOFAD are caused by mutations that influence the production of Aβ. However the majority of AD cases have a more complex aetiology and cannot be directly linked to abnormal Aβ production. The more common form of AD is not pathologically distinct from EOFAD, so a major hypothesis is that LOAD is caused by an imbalance of Aβ concentrations in the brain, which results from alternative molecular mechanisms from EOFAD (see Section 1.6).

The main pathways that have been targeted by researchers in the search for alternative mechanisms that influence Aβ deposition, are Aβ degradation and clearance from the brain. Soluble, monomeric Aβ can cross the blood brain barrier (BBB) through interactions with two receptors: receptor for advanced glycation end products (RAGE, influx) and low-density lipoprotein receptor related protein-1 (LRP-1, efflux) for review see Deane *et al.* and Citron, M. [186, 187]. Inhibition of RAGE and up-regulation of LRP-1 are potential therapeutic strategies that are being explored to decrease uptake and increase removal of Aβ from the brain into the peripheral circulation, where it is largely bound to other proteins so cannot form toxic aggregates. The long term effects of continued Aβ increases in the blood is unknown however.

In addition to receptor mediated transport across the BBB, brain Aβ levels can also be altered by the action of Aβ peptidases and several enzymes that are known to degrade Aβ. These include, but are not limited too, neprilysin (*MME*), insulin degrading enzyme (*IDE),* angiotensin converting enzyme (*ACE*), plasmin (*PLG*), endothelin converting enzymes (*ECE-1* and *ECE-2*) and matrix metallopeptidases (*MMP-2, MMP-3* and *MMP-9*) [188-193]. These enzymes and factors that regulate the expression and activity of these enzymes have been the target of both genetic and biochemical research, in an effort to identify molecular mechanisms that can be exploited therapeutically. Whilst each of the above peptidases have evidence implicating them as able to degrade Aβ, their impact on AD pathogenesis is less clear. Furthermore, whilst neprilysin is probably the most established Aβ peptidase, genetic evidence for association with AD is more compelling for the genes encoding ACE and IDE. The association of SNPs in genes encoding one of these enzymes (IDE), with plasma Aβ levels and AD is explored further in Chapter 3.

One final area of research focused on Aβ clearance from the brain is that of innate immunity. As mentioned previously, neuritic (or dense) amyloid plaques are associated with microglial activation, indicating that an inflammatory response is a component of the disease. Recent studies suggest that phagocytosis of Aβ by macrophages may be defective in AD patients [194, 195]. Studies carried out in mice also show that administration of Aβ antibodies stimulate clearance of Aβ by activation of micoglia [196]. This work is based upon research published in 1999 and 2000 demonstrating that vaccination of AD mouse models with Aβ1-42 resulted

in reduced deposition of Aβ [197] and improvement in memory phenotypes [198]. A similar vaccination protocol as a therapeutic treatment for AD reached clinical trials in 2002 [199, 200]. Unfortunately the trials were cancelled due to unexpected toxicity caused by a neuroinflammatory response (meningo-encephalitis) in some patients. The field continues to pursue this avenue of research in order to develop a treatment for AD [201, 202].

1.10 Measurement of Aβ in human samples

Aβ can be measured in samples of CSF and plasma and has also been measured in homogenates of human brain [66]. The availability of plasma is much greater than that of CSF so, more work has been carried out using this source. In plasma, contrary to what is found in plaques, Aβ42 is not the most abundant form of the peptide. In published studies measuring plasma Aβ40 and Aβ42, plasma concentrations of Aβ40 range from 2 to 10 times as high as those for Aβ42 [95, 203].

In previous studies of EOFAD families, it was found that all of the genetic changes that cause AD alter APP processing in a way that is evident in plasma [95]. In 2001, Ertekin-Taner *et al*. demonstrated that in a study of 15 typical LOAD pedigrees, plasma Aβ40 and Aβ42 showed highly significant heritability ($p<0.0001$) of 41% and 56% respectively [204]. In additional studies by the Younkin laboratory, plasma Aβ concentrations in extended LOAD pedigrees were also used to identify a QTL on chromosome 10 [62] and subsequent association with SNP's at several genes proximal to the linkage peak [205-207].

Several studies have been published over the past 14 years demonstrating the utility of plasma Aβ measures in genetic and epidemiological studies of AD. Many of these studies made use of a sandwich ELISA system developed by N .Suzuki in the Younkin lab [94].

In this thesis, the same sandwich ELISA system has been used to measure plasma Aβ in population isolates in order to identify genetic loci that influence this trait. This is a novel approach, since the study subjects have not been selected based on *a priori* knowledge of cognitive deficits or individuals with high plasma Aβ levels.

1.11 Genetic factors that influence LOAD

Most cases of early onset AD segregate in families and are caused by mutations in three genes (*APP, PSEN1* and *PSEN2*) that show an autosomal dominant pattern of inheritance. However, the more common late-onset AD (LOAD) disorder does not exhibit an autosomal dominant pattern of inheritance but is highly heritable with an estimated heritability of 58-79%, based upon twin studies [208]. Thus, understanding genetic factors that influence AD is clearly pertinent to understanding the disease, predicting disease risk and development of targeted therapeutic strategies.

Several studies have investigated the *APP*, *PSEN1* and *PSEN2* loci for single nucleotide polymorphisms (SNPs) that might influence LOAD, with largely negative findings [31, 209]. However some evidence suggests a possible role for SNPs at these loci that may influence gene expression [210-213] or interaction with another locus such as *APOE* [214, 215].

Following on from the success of linkage studies in EOFAD pedigrees, several groups undertook a similar approach using LOAD pedigrees and in 1991 linkage to chromosome 19 was observed [216]. In 1993, apolipoprotein E (*APOE)* on chromosome 19 was determined to be a risk locus for LOAD based upon co-incident studies that identified association of *APOE* alleles with LOAD [217, 218] and the finding that ApoE and Aβ are binding partners [219-221].

Whilst *APOE* genotype is now a well established risk factor for LOAD, it does not explain all of the genetic variance of LOAD. *APOE* risk alleles are neither necessary nor sufficient for development of LOAD, consequently additional LOAD linkage studies were undertaken in an effort to identify further genetic risk loci. These linkage studies were followed by candidate gene studies, often analysing variation in genes under the identified linkage peaks.

Bertram *et al.* created a database called Alzgene in order to provide convenient access to results from the large numbers of LOAD genetic association studies [31]. The primary purposes of AlzGene are to record the results of published AD association studies and to perform meta-analyses.

At the time of writing there are data from 2916 polymorphisms collected from 1380 studies of 666 genes in the AlzGene database. These numbers illustrate the substantial effort by researchers over the past 15-20 years in an attempt to identify LOAD susceptibility loci. Despite this effort, *APOE* remains the “top gene” on a list of genes with nominally significant meta analysis p-values, ranked based upon HUGENet criteria and compiled by AlzGene [222, 223].

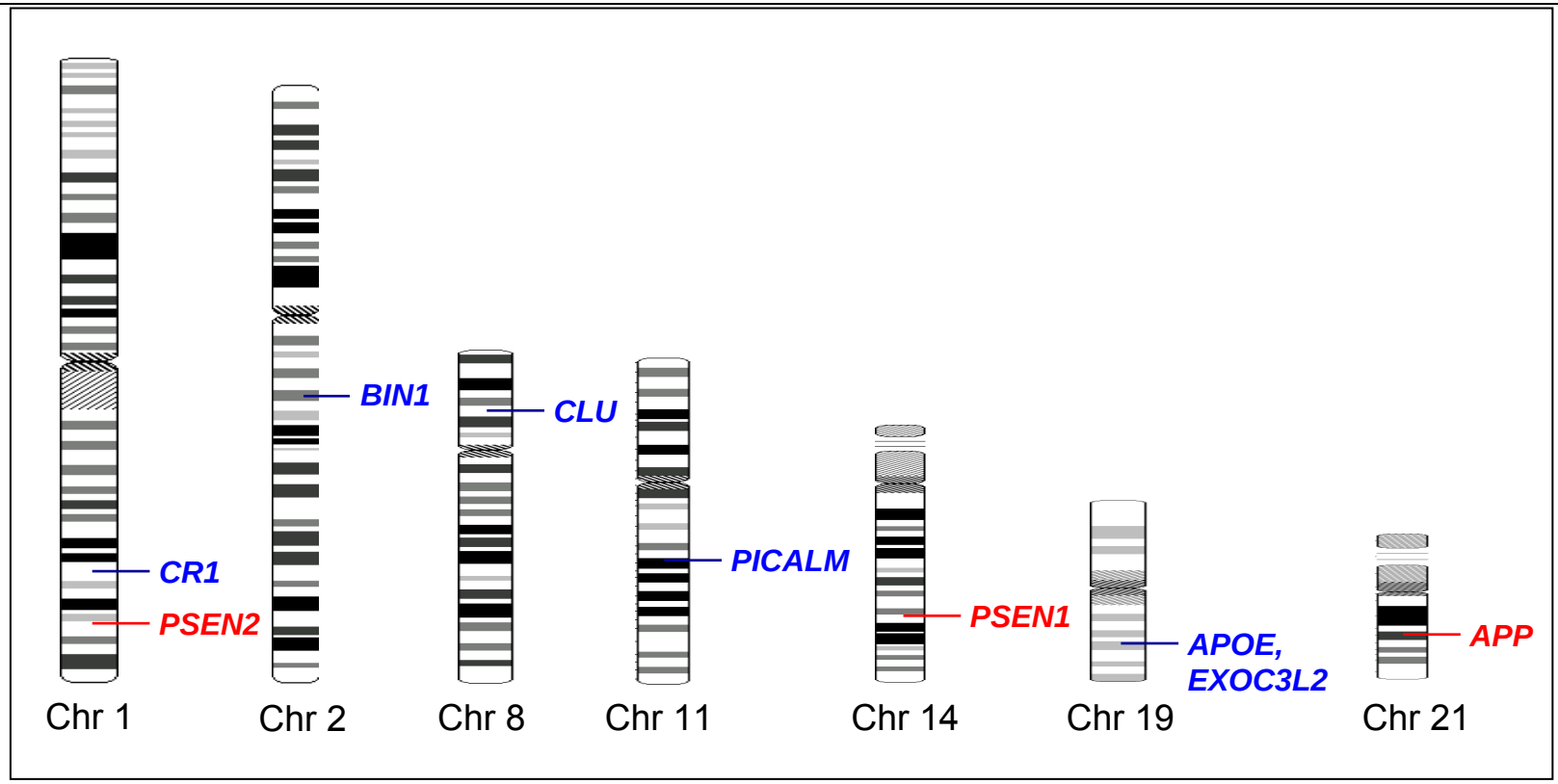

Figure 1.6 Ideogram representations of 6 human chromosomes with genes known to influence AD. Genes identified that are known to cause EOFAD are shown in red and genes identified that are known to influence risk for LOAD are shown in blue. Ideograms are adapted from the Alzgene website [31]. Chr = chromosome.

Furthermore, *APOE* remained, until recently (and may still be considered by some), the only gene to be irrefutably accepted as a LOAD susceptibility gene, due in large part to inconsistent replication of other genes and variants in additional populations. The effect size of *APOE* is also much larger than any subsequently identified loci, although some studies initially report inflated odds ratios, a phenomenon referred to as 'the winner's curse' [224]. It is likely that if additional 'common' LOAD loci with an effect size equal to or larger than *APOE* do exist, they would have been identified already. Consequently the remaining genetic risk for LOAD is now believed to be due to multiple gene variants with modest effect size or rare gene variants with moderate to large effect sizes, since the latter are not efficiently screened by genetic association studies using common variants [225]

1.11.1 *APOE*

APOE is located on chromosome 19q13.2, flanked by two genes: *TOMM40* 3' and *APOC1* 5'. A two SNP Haplotype (rs7412 and rs429358) tags three common *APOE* alleles, namely *APOE E2*, *E3* and *E4* [226]. The *APOE E4* allele confers increased risk for disease with an OR of ~4.0 for the *E4* versus the *E3* allele or an OR ~ 15.0 for *E4/E4* homozygotes versus *E3/E3*homozygotes (Alzgene meta analysis[31]). The *APOE E2*allele is believed to confer protection from AD but with a weaker effect than *APOE E4* [227]. In addition to the effects of the *APOE* locus on disease risk, it has also been reported that the *E4* allele can influence the age of onset of the disease, where *E4* carriers have an earlier age at onset of the disease when compared with non-carriers [217].

Several studies have been published recently postulating that genes lying proximal to *APOE* may contribute to or account for the genetic association at the *APOE* region or independently modify risk for LOAD [228]. Whilst this remains controversial, it is possible that disease modifying variants in multiple closely linked genes might influence the disease, that might help to explain the genetic heterogeneity observed at other LOAD loci.

APOE encodes the apolipoprotein E, which is the main cholesterol chaperone in the brain. It has been hypothesized that one way in which the APOE protein might influence risk for AD is through altered cholesterol trafficking. Cholesterol is a key component of lipid rafts and as mentioned in Section 1.6, cholesterol concentration in lipid rafts may influence APP trafficking/processing and consequently risk for AD. It is thought that the APOE ε4 protein isoform, encoded by the *E4* allele, binds cholesterol less efficiently than the *E3* and *E2* isoforms and that the resulting effects on cholesterol trafficking may impact disease [229]. The relationship between cholesterol, ApoE and neurodegeneration is, however, likely to be complex and may be modified by many other genetic and environmental factors (see recent reviews [230, 231].

1.11.2 Novel LOAD loci identified by GWAS

Significant progress has been made in identifying LOAD susceptibility genes over the past two to three years with the advent of genome-wide association studies (GWAS). Studies with a sample size of >100 are summarized in Table 1.4. GWAS can be highly informative since they interrogate the whole genome with high density

coverage, and they provide an unbiased approach to screen for SNPs in novel trait-associated genes. The majority of LOAD GWAS published to date (Table 1.4) genotyped between 300,000 and 500,000 SNPs using microarray technology. To illustrate how quickly this technology has evolved, the most recent platforms are offering microarrays that genotype over 1 million SNPs and there are plans for platforms offering 5 million SNP genotypes (www.illumina.com, November 2010). GWAS also present challenges, however, in the form of multiple testing. Corrections for multiple tests must be performed which result in a low p-value threshold (approx 1.7E-07 for 300,000 tests) to pass significance at an α of <0.05. Many of the early LOAD GWAS studies were unable to identify loci that passed the stringent criteria for multiple test corrections with the exception of SNPs on chromosome 19 proximal to the *APOE* locus (Table 1.4). Genes that were implicated in these earlier studies were not replicated consistently, similar to the results observed for candidate gene studies, likely due lack of statistical power in either the initial or subsequent follow-up studies. [232-234]. However as sample sizes increased more promising results (and smaller p-values) emerged.

Many of the most recent studies are the product of large collaborative efforts between multiple research groups, for example the Genetic and Environmental risk in Alzheimer's disease 1 consortium (GERAD1) and the European Alzheimer's Disease Initiative stage 1 (EADI1) consortium which published LOAD GWAS data in 2009.

GWAS Authors	Reference	Design	Population(s)	No SNPs	No AD GWAS (Follow-up)	No CON GWAS (Follow-up)	Highlighted genes *
Grupe et al 2007	[235]	Case-control	USA & UK	17,343	380 (1428)	396 (1666)	*APOE*
Coon et al 2007, Reiman et al 2007	[232, 236]	Case-control	USA, Netherlands **(a)**	502,627	446 (415)	290 (260)	*APOE, GAB2*
Li et al 2008	[237]	Case-control	Canada & UK	469,438	753 (418)	736 (249)	*APOE*
Abraham et al 2008	[238]	Case-control	UK	561,494	1082 (n.a)	1239(1400)	*APOE*
Bertram et al 2008	[239]	Family based	USA	484,522	941 (1797)	404 (838)	*APOE, ATXN1, CD33, GWA_14q31*
Beecham et al 2009	[240]	Case-control	USA **(d)**	532,000	492 (238)	498 (220)	*APOE, FAM113B*
Carrasquillo et al 2009	[241]	Case-control	USA **(b)**	313,504	844 (1547)	1255 (1209)	*APOE, PCDH11X*
Lambert et al 2009 (EADI1)	**[242]**	Case-control	Europe **(c)**	~540,000	2032 (3978)	5328 (3297)	*APOE, CLU, CR1*
Harold et al 2009 (GERAD1)	**[243]**	Case-control	USA & Europe **(b,c)**	~610,000	3941 (2023)	7848 (2340)	*APOE, CLU, PICALM*
Seshadri et al 2010	[228]	Case-control	USA & Europe **(a,b,c)**	~2,540,000 #	3006 (6505)	22,604 (13,532)	*APOE, BIN1, CLU, EXOC3L2, PICALM*
Naj et al 2010	[244]	Case-control	USA & Europe **(a, d)**	483,399	931 (1338)	1104 (2003)	*APOE, MTHFD1L*
Lee et at al 2010	[245]	Case-control	Caribbean Hispanic	658,610	549 (n.a)	544 (n.a)	-

Table 1.4. LOAD GWAS studies.Table modified from Bertram *et al* [246]. Summary of published GWAS studies (with an n>100) where the phenotype tested is clinical or pathological diagnosis of LOAD. *Genes highlighted by the indicated study to have achieved study-wide/genome-wide significant association: the criteria used to define 'genome-wide' significance within a study is not standardized and may vary between studies. Overlap in subjects used between studies is indicated under the populations field (**a,b,c or d**). # Seshadri et al report imputed SNP data.

In 2009 the largest LOAD GWAS, at the time of publication, were presented. Carrasquillo et al identified genome-wide significant association for a SNP within the *PCDH11X* gene on ChrXq21.3 [241] in a two-stage study totalling 2391 LOAD cases and 2464 elderly controls (Table 1.4). This was the first LOAD GWAS to unequivocally identify genome-wide significant association, independent from *APOE*, in an un-stratified dataset according to the strictest correction for multiple tests (Bonferroni). The data obtained from this GWAS are used later in this thesis to perform convergent phenotype analysis (Chapter 3) and as part of a candidate gene study (Chapter 4).

Later in 2009 Harold *et al* and Lambert *et al* [242, 243] independently published two large GWAS which combined data from several research labs (the Harold *et al* study includes the data from the Carrasquillo et al GWAS). The two studies identified genome-wide significant association at the *APOE* locus and an additional three loci, remarkably one of which was common to both reports. Two SNPs were identified by Harold *et al*: rs11136000 on chromosome 8p21-p12 within an intron of the *CLU* gene (OR = 0.86, p = 8.5E-10) and rs3851179 on chromosome 11q14 ~90kb 5' of *PICALM* (OR = 0.86, p = 1.3E-09). In the Lambert *et al* study rs11136000 within *CLU* also significantly associated with LOAD (OR 0.86, p = 7.5E-09) and the authors identified association for a second SNP rs6656401 (OR = 1.21, p = 3.7E-09) within an intron of the *CR1* gene on chromosome 1q32. Over the past year several reports of replication of the association observed at these three loci provide evidence in support of the SNPs proximal to *CLU, PICALM* and *CR1* in LOAD pathogenicity [228, 245, 247-250].

In 2010 three LOAD GWAS have been reported thus far [228, 244, 245]. Seshadri et al present data from several combined LOAD GWAS (both novel and previously published GWAS) where SNP imputation allowed for the combination of data obtained from multiple genotyping platforms and created the largest LOAD GWAS to date. The previously reported association at the *CLU* and *PICALM* loci were confirmed and an additional two novel loci were identified that achieved genome-wide significant association with LOAD. rs744373 on chromosome 2q14, ~30kb 3' of *BIN1* (OR = 1.13, p = 1.59E-11) and rs597668 on chromosome 19q13.32 ~7kb 5' of *EXOC3L2* (OR = 1.18, p = 6.45E-09). This locus is ~300kb distal of *APOE,* the authors indicate that the association reported for rs597668 is independent of *APOE*.

The Seshadri *et al* study overlaps with another LOAD GWAS, also published in 2010, reported by Naj *et al* [244]. In this study genome-wide significant association was identified for a novel LOAD locus, rs11754661 within the *MTHFD11* gene on chromosome 6q25.1 (OR = 2.03 , p = 4.7E-08). The authors reported significant replication of this finding in an independent case-control series, however this finding is so novel that additional replication has not yet been reported.

In summary, to date, more than 10 GWAS have identified several SNPs with strong association with LOAD. Multiple studies have reported loci which achieved study-wide significance, of which seven (*CLU, CR1, PICALM, BIN1, EXOC3L2*, *PCDH11X* and *MTHFD11*) maintain significance after conservative bonferonni correction for multiple tests.

The first five genes listed in the previous paragraph are also ranked from positions 2-6 in the AlzGene 'tophits' list (at the time of writing). As mentioned previously *APOE* has consistently occupied position 1 on this list which ranks genes on the basis of nominally significant meta analysis p-values [31]. The remaining genes or loci that make up the 'top ten' AlzGene genes are *SORL1* [251], a locus on Chr14q32.13 [235], *TNK1* [235]and *IL8* [252]. These genes were initially identified as LOAD susceptibility genes through GWAS or candidate gene studies and they represent additional LOAD loci which have strong evidence for a role in disease aetiology.

The associations reported here are of modest effect size (OR ~ 0.8 – 1.3) with the exception of the most recent GWAS locus (MTHFD11) identified with an OR ~2. Although these effect sizes are small they may still, in principal, be useful for identifying therapeutic targets and the at risk population. Of particular relevance is the emergence of genes involved in common pathways, for example CLU (clusterin) also known as APOJ (apolipoprotein J) is, like ApoE, a cholesterol transport protein and is the second most abundant of the cholesterol chaperones in the CNS. Several other genes involved in cholesterol metabolism have been implicated in LOAD [230]. The implications of variation in these genes on LOAD pathogenicity is discussed in more detail in each of the respective publications and summarized in Figure 1.9 at the end of this chapter.

1.12 Approaches for studying complex genetic diseases.

Over the past decade advances have been made in technological and methodical approaches for interrogating the human genome for markers of human disease. In 2003 the International Hapmap project was launched in which the aim was to map common variation in the human genome which can be used as a tool by researchers to identify genes involved in human disease [253]. For example the data collected by the Hapmap project was used to identify SNPs that tag common haplotypes in certain populations. This data facilitated the development of panels of 300,000 to 500,000 SNPs that constitute the main component of genome-wide genotyping platforms used for GWAS. More recently an initiative to identify less common variation (as low as 0.1%) in the human genome was launched called the 1,000 Genomes Project. This project involves sequencing the genomes of over 2000 subjects from many ethnic backgrounds in an effort to identify all genetic variation in the human genome with a frequency of 1% or greater and to catalogue identified rare variants (frequency as low as 0.1%). Currently the project has sampled 500 people from each of five ancestral backgrounds (European, West African, East Asian, America's and South Asian) [254]. This is the first large scale study to explore new sequencing technology referred to as 'next-generation sequencing'. The ability to sequence the entire genome or exome of study subjects is a powerful new method, considered to be the next phase in the exploration of the human genome consistent with the overall aim of identifying variation that influences human phenotypes and disease. This novel approach has only recently become economically feasible and it is still costly for sequencing large numbers of subjects.

It is anticipated that over the next 2-3 years as the cost of this sequencing technology decreases there will be a vast amount of data available for the exploration of the genetic causes of human disease. Given the novelty of this technology the approach was not explored in this thesis but will be useful for future studies.

With respect to AD and other complex diseases such as diabetes, it has become apparent that large numbers of cases and controls are necessary for identifying association to common variants in novel disease loci using a case-control approach. Since large case-control series (many thousands of subjects) are not readily available to most researchers, other approaches have been developed to improve the success rate, such as studies aiming to identify quantitative trait loci (QTL) or disease susceptibility loci. These include the use of endophenotypes, isolated populations and phenotypic convergence analysis, all of which are explored in this thesis.

1.12.1 Use of endophenotypes and convergence approaches.

The molecular mechanisms that lead to a disease outcome in complex genetic diseases such as AD, schizophrenia and type 2 diabetes mellitus are, by nature, complex and sometimes difficult to explore using a categorical disease outcome phenotype. To this end, the use of intermediate disease related phenotypes or 'endophenotypes' can be useful for identifying susceptibility genes. This approach is based on the hypothesis that phenotypes or traits that represent certain aspects of a disease have a simpler genetic aetiology than the full disease trait. This approach has been used to improve our understanding of genetic factors that influence complex genetic disorders such as type-2-diabetes [255] and schizophrenia [256, 257].

Likewise several relevant endophenotypes for exploring the genetics of AD have also been reported. These include plasma concentrations of Aβ discussed previously and CSF levels of tau and Aβ as well as neuroimaging measures, cognitive traits and gene expression levels measured from brain tissue samples.

In this thesis plasma Aβ levels (concentration) are measured and explored as an AD endophenotype in genome-wide linkage and association studies in an effort to identify novel LOAD candidate genes. Furthermore this data is combined with that obtained from parallel studies of gene expression levels and AD disease phenotype in a phenotype convergence analysis described in chapter 3.

1.12.2 Utility of isolated populations.

Complex genetic diseases such as AD are heterogeneous and likely to be influenced by multiple genes with small effect sizes, which can make the identification of disease susceptibility loci challenging. One approach to reduce genetic and environmental heterogeneity in a study is to use isolated populations.

Population isolates are defined as populations arising from a small number of people (founders) that have experienced little admixture for many hundreds, or in some cases thousands, of years. In these populations genetic drift has resulted in reduced allelic heterogeneity and extended linkage disequilibrium (LD), making them an attractive resource for gene mapping of complex diseases and traits [258]. In population isolates it is likely that there are fewer genetic factors that influence quantitative traits or diseases. The effect size of these loci may be larger than in outbred populations and therefore more easily identified and mapped. Based on this

hypothesis several research studies have been conducted on isolated populations throughout the world to identify susceptibility genes for complex disease and traits. This approach proved successful for rare recessive disorders [259], Mendelian disorders and more recently for more common complex traits such as type-II-diabetes, schizophrenia and other diseases and disorders [260-263]

In this thesis isolated populations sampled from two Croatian islands (Vis and Korcula) are used to identify genetic loci that influence plasma Aβ levels in an effort to identify novel candidate LOAD loci. These populations have been well characterized and exhibit many features associated with population isolates [264-268]. However these populations also exhibit enough similarity to outbred populations of European origin, for example in sharing of common variants influencing traits, for QTL to be identified in these studies that are also relevant in larger outbred populations [269]. Over 300 anthropometric and biochemical traits have been measured in the study samples, which have also been genotyped for over 300,000 SNPs, making genome-wide association analysis possible for each of the measured traits [270, 271]. The evaluation of serum uric acid concentrations in these populations led to the identification a novel locus, SLC2A9, which was shown to be a novel urate transporter, influencing genetic susceptibility to gout as well as variation in this quantitative trait. This finding was confirmed in additional populations illustrating the utility of the approach for identifying relevant and novel QTL [272, 273].

Whilst the populations sampled from Vis and Korcula, represent excellent study populations for mapping QTL, they were not ascertained on the basis of *a priori*

knowledge of cognitive traits, AD diagnosis or levels of segregating plasma Aβ levels.

1.13 Mitochondrial role in AD

Several studies have outlined how mitochondrial dysfunction can impact neuronal cell function and how this may influence risk of neurodegenerative diseases such as AD.

1.13.1 Mitochondrial theory of aging

The primary role of mitochondria within the cell is the production of ATP through the transfer of electrons from donors, such as NADH, to a series of electron acceptors, such as oxygen, in which electron transfer is coupled to hydrogen ion transport across a membrane to create a gradient that drives ATP formation (oxidative phosphorylation, OXPHOS) (Figure 1.8). In addition to OXPHOS, the mitochondrial compartment is the major site for Krebs cycle reactions, β-oxidation of fatty acids, calcium homeostasis, lactate metabolism and the biosynthesis of iron-sulphur proteins, nucleotides and certain amino acids and phospholipids

Mitochondria are also a key component of an apoptotic signalling cascade in which, following opening of the permeability transition pore (mPTP), a specific pore in the outer mitochondrial membrane, pro-apoptotic signalling proteins (e.g. cytochrome C) are released into the cytoplasm, triggering a cascade of reactions including caspase activation, degradation of DNA and ultimately apoptotic cell death. Apoptosis can be triggered by a number of factors, among which is damage to

proteins, membranes and DNA through interaction with free radicals (reactive oxygen species). The main source of reactive oxygen species (ROS) within a cell is electron leakage from the electron transport chain (Fig 1.8) [274].

One way in which the mitochondrial genome could act as a redox sensor is through ROS-mediated damage to mitochondrial DNA (mtDNA). The mitochondrial theory of aging postulates that mtDNA accumulates mutations with time and that this in turn increases mitochondrial ROS production, increasing age-related cellular dysfunction and eventually triggering apoptosis [275, 276].

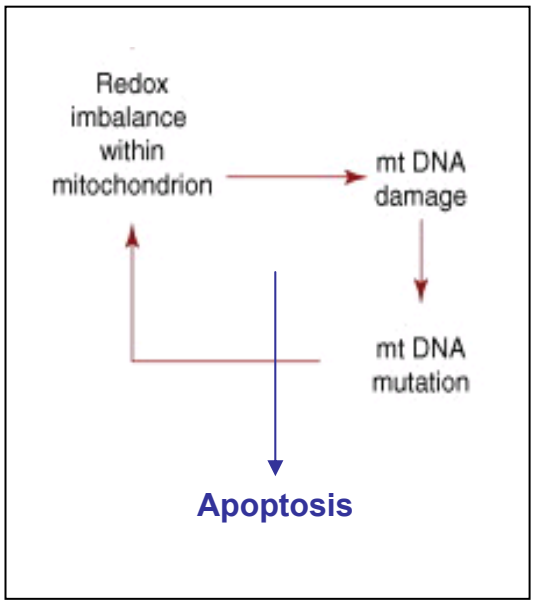

Figure 1.7 Simple interpretation of the mitochondrial theory of aging. This represents how the mitochondrial genome could act as a redox sensor. A feedback loop maintains a vicious cycle, resulting in cellular dysfunction and eventually triggering apoptosis. Adapted from Wright et al [277]

1.13.2 The mitochondrial genome

Each cell contains multiple (10^3-10^5) mitochondria and each mitochondrion contains several (5-10) copies of the mitochondrial genome. The mitochondrial genome is particularly vulnerable to ROS induced DNA damage due to its close proximity to the main cellular source of ROS (the electron transport chain), lack of protection by associated proteins such as histones, lack of DNA repair pathways (e.g. nucleotide excision repair) and the ratio of coding to non-coding DNA [278]. Estimates place the mutation rate of mtDNA in the brain up to 10 times as high as for nDNA [279]

The mitochondrial genome is a closed circular double stranded DNA molecule of approx 16kb in size. It contains 37 genes which encode 13 proteins essential for the electron transport chain, 2 ribosomal RNA's and 22 tRNA genes. The mitochondrial genome has very little non-coding sequence with the only notable region being the D-loop locus which contains the heavy (H) strand origin of replication and both H and light (L) strand transcriptional promoters. This region is also often referred to as the 'hypervariable' region due to the fact that it is highly polymorphic when compared with the reference coding sequence of the mitochondrial genome [280].

There is only one known DNA polymerase that is present in mitochondria, known as polymerase gamma (Pol-γ). This polymerase is responsible for mtDNA replication, proof-reading and DNA repair functions. It is encoded by a nuclear polymerase gamma (*POLG*) gene located on chr15q25 and following transcription and translation in the cytoplasm, it is transported into the mitochondria. Mouse mutants with polymerase gamma (*Polg*) lacking 3'-5' exonuclease activity (proof reading deficient) show a two orders of magnitude increase in single nucleotide sequence

variants compared with control mice and exhibit premature aging phenotypes [281, 282]

In humans and a wide range of other species, the mitochondrial DNA of many post-mitotic tissues accumulates DNA mutations over the course of a lifetime [278]. Increasing the frequency of acquired mtDNA mutations is associated with increased evidence of aging and cell death (as in the *Polg* "mutator" mouse) [282, 283], while decreasing their frequency reduces the signs of aging and cell death [284]. In post-mitotic cells such as neurons, this process may have a significant impact on cell function and survival.

1.13.3 mtDNA damage and mutations in AD

Several lines of evidence indicate a role for mitochondrial dysfunction in Alzheimer's disease (for a comprehensive review see de Moura *et al* [285]). Of particular interest is the possibility that mitochondrial mutations can either cause or influence risk for AD and other neurodegenerative diseases. This might occur through inherited common mutations, acquired somatic mutations that have reached a high frequency over time through replication, or by a cumulative burden of somatic mutations with individual low frequencies that collectively result in impaired mitochondrial function (mitochondrial burden) [286].

Previously studies have implicated common mitochondrial mutations and mitochondrial haplogroups in age related neurodegenerative diseases such as AD and PD [287-291]. Additionally some studies report association of mtDNA deletions with Parkinson's disease [288] and others have investigated a common 4997bp

deletion (known as the Kearns-Sayre syndrome deletion) with respect to these diseases [292-294]. More recently studies have focused on the investigation of mitochondrial mutational burden as a possible risk factor for development of neurodegenerative diseases [287, 295, 296]. However replication of these reported associations has been inconsistent [297, 298] and the role of mtDNA mutations in Alzheimer's disease remains controversial.

Another line of inquiry into the possible role of oxidative DNA damage and risk for diseases such as AD is the evaluation of mtDNA damage lesions. Several studies have identified increased markers for mtDNA damage in the brains of AD subjects when compared with age matched controls, [279, 290]. However it is currently unclear as to whether this is a feature common to many neurodegenerative diseases or a possible disease process specific to AD.

In this thesis various methods of measuring mitochondrial DNA lesions are explored and investigated for association with neurodegenerative diseases.

1.13.4 Complex IV and AD

In order to produce ATP, the mitochondria transports electrons through a system of transmembrane protein complexes creating a proton gradient across the inner-mitochondrial membrane which drives the production of ATP (Fig 1.8). All of these protein complexes, except for complex II, rely on subunits encoded by both the nuclear and mitochondrial genome. Complex II relies entirely on nuclear encoded proteins.

Several studies have reported deficiencies in energy metabolism associated with AD and other neurodegenerative diseases [299-301]. One of the most frequently associated changes in energy metabolism associated with AD is that of complex IV (cytochromc c oxidase) deficiency [302, 303]. Furthermore, evidence of from cybrid studies indicates that altered mitochondrial metabolism is associated with Aβ toxicity [304] indicating a link between mitochondrial dysfunction and AD pathogenicity. Additionally, evidence from some studies indicates that APP may be partly responsible for the observed complex IV deficiencies in AD [305]. The association between APP, Aβ and mitochondrial dysfunction in AD is reviewed briefly below.

1.13.5 APP, Aβ and mitochondria

As shown in Figure 1.3, data from some studies suggests that APP can be targeted to the mitochondrial membranes where it is incompletely translocated across the mitochondrial membranes due to a transmembrane arrest sequence. This can result in the blocking of the TOM channel (translocase of the outer membrane of mitochondria) and thus inhibit the translocation of other proteins and molecules into the mitochondria [139, 305] which can disrupt mitochondrial function. Furthermore data from mouse studies implicates accumulation of APP at the mitochondrial membrane that can result in impaired energy metabolism (the central role of the mitochondria) and consequently mitochondrial dysfunction [139].

In addition to APP, the Aβ peptide can also be localized to the mitochondria as first indicated by Lustbader *et al* [306]. Aβ found within the mitochondria and can

interact with a number of mitochondrial proteins causing downstream mitochondrial dysfunction [306]. This indicates an additional mechanism by which Aβ can exert toxicity. For example, Aβ can bind to The Aβ-binding alcohol dehydrogenase (ABAD) which results in up regulation of cytochrome c release and consequently increase apoptosis signalling [306]. More recently cyclophin D (CypD) was also found to interact with Aβ within the mitochondria. CypD is located within the mitochondrial matrix but can be transported to the inner mitochondria membrane where it is thought to participate in the formation of the mitochondrial permeability transition pore(mPTP). As mentioned previously (1.13.1) the formation of the mPTP allows the release of pro-apoptotic signalling proteins. The combined interaction of Aβ with ABAD and CypD therefore increases pro-apoptotic signalling and may represent a relatively novel molecular mechanism by which Aβ exerts neuronal toxicity.

The evidence presented on the previous pages clearly indicate a role for mitochondrial dysfunction in AD although the mechanisms by which this can occur appear to be multiple and likely heterogeneous. Furthermore it is probable that cellular Aβ concentrations and mitochondrial dysfunction are connected possibly in a vicious feedback loop whereby increased Aβ causes increased mitochondrial dysfunction which in turn may increase Aβ production [285, 304].

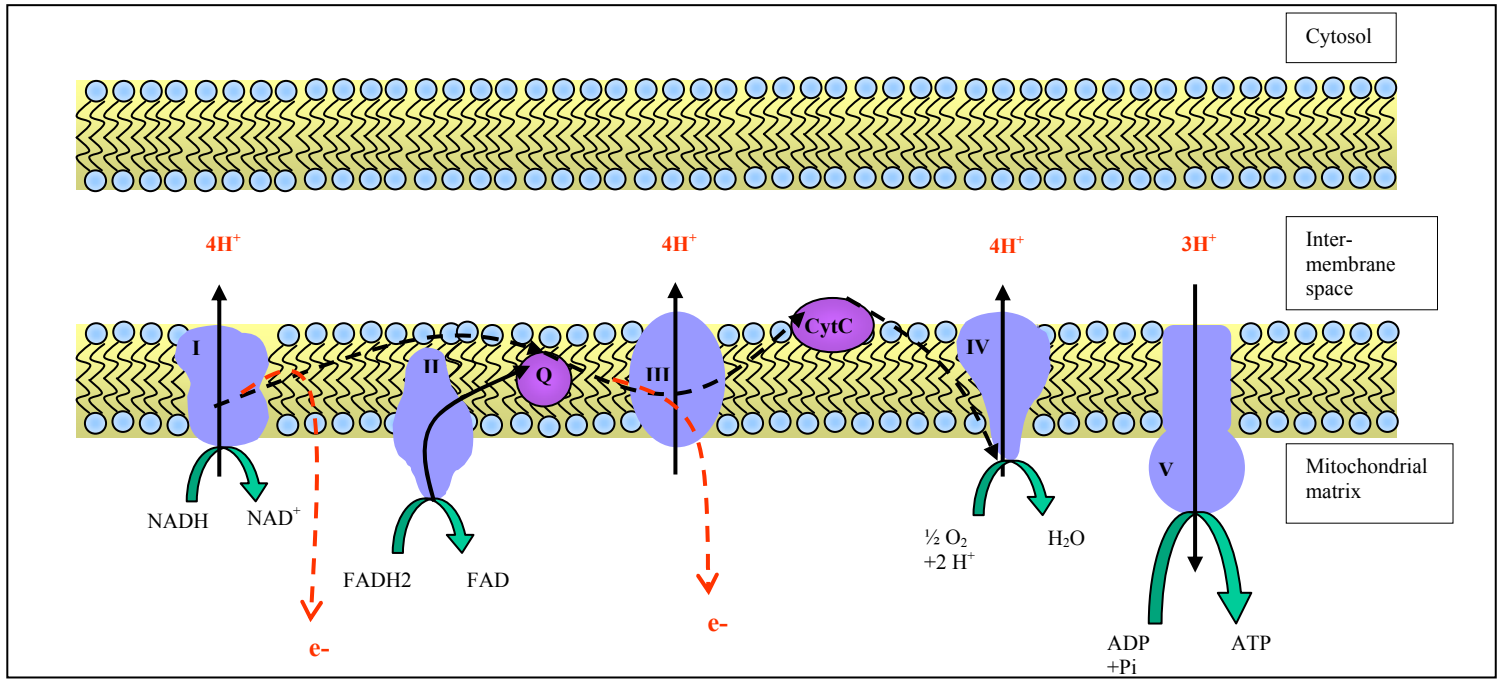

Fig 1.8 The mitochondrial electron transport chain: The mitochondrial electron transport chain facilitates the production of ATP by transport of electrons through a system of protein complexes (I-IV) to generate a proton gradient that 'powers' the ATP synthase (V). Here, the solid black arrows illustrate the flow of hydrogen ions trough the protein complexes, whilst the black dashed lines represent the normal flow of electrons through the transport chain ending in the production of H_2O. The red dashed lines illustrate electron leakage at the two most common sites in the transport chain: complex I and complex III/Q.

These studies emphasise the importance of the exploration of multiple aspects of the disease in order to elucidate the molecular mechanisms involved in AD.

1.14 Environmental influences on risk for LOAD

In addition to a genetic component for risk for LOAD, there is also a substantial influence of environmental effects on disease risk and/or age at onset of the disease. In fact age is the major risk factor for this disease where its influence on risk for the disease is greater than any one genetic risk factor, including *APOE*.

It has been estimated that the risk for developing AD in the general population can almost double every five years after the age of 65 and up 50% of people over the age of 85 may have AD although these estimates may be slightly inflated [307].

In addition to age, a number of other environmental factors have been reported to modify risk for the disease; some of these include diet [308, 309], head injury [310], gender [307] and years of education [307]. It is believed that a complex interaction between environmental and genetic risk factors for the disease exists which modifies risk and age of onset for the disease [311].

Similar to the AlzGene database, the Alzheimer research forum (Alzforum) supports an AlzRisk database which

> "aims to provide a comprehensive, unbiased, centralized, publically available and regularly updated collection of epidemiological reports that evaluate environmental (ie non-genetic) risk factors for Alzheimer disease (AD) in well defined study cohorts" (www.alzrisk.org, Dec 2010) [312].

At the time of writing, the AlzRisk database highlights studies evaluating the influence of Diabetes Mellitus, Inflammatory biomarkers, Physical activity and Vitamin E on risk for developing AD.

It is still unclear how much of an influence non-genetic factors have on risk for AD but the general consensus is that a healthy lifestyle can likely reduce the risk for many diseases associated with old age including AD.

1.15 Outline of the Project

The aims of the projects in this thesis are to identify novel genetic risk factors for LOAD and to develop novel approaches for investigating the genetic risk for LOAD. The studies presented here are not limited to the evaluation of nuclear genetic risk factors and include the investigation of the mitochondrial genome for markers associated with disease.

Based upon the evidence from previous LOAD GWAS it has become apparent that large numbers of cases and controls are needed to identify novel LOAD genes with moderate effect size. In this thesis several alternative approaches are combined and explored in order to identify novel LOAD loci without the need for infinitely large sample size. An AD related endophenotype, plasma Aβ, was used to identify novel loci that influence this trait in isolated populations for which genome wide microsatellite (linkage) and/or SNP GWAS data was available. It is hypothesised that loci that influence plasma levels of Aβ may be excellent LOAD candidate loci because Aβ is a key peptide in AD pathogenesis. The utility of plasma Aβ as a LOAD endophenotype in genetic studies has been demonstrated previously.

Additionally isolated populations are thought to be a useful tool for studying complex genetic diseases where genetic heterogeneity in outbred populations can result in reduced power for detecting loci that influence complex traits or diseases.

The results of the study described above generate a database of p-values that can be interrogated in phenotype convergence analysis. It is well established that LOAD is likely influenced by multiple genes with individual moderate effects that are difficult to detect. However evidence for association of a locus with multiple disease related traits can strengthen the plausibility for a role in the targeted disease. This approach was explored for a well known LOAD candidate gene *IDE*, using the data from a LOAD GWAS, the plasma Aβ GWAS described in this thesis and measured expression levels of *IDE* in the brain. In the future this approach can be applied to additional candidate genes.

Finally, much of the focus of genetic studies of LOAD is concentrated on identifying variation in nuclear chromosomes that influences risk for disease. However the variation at the mitochondrial genome is emerging as an important factor that can influence neurodegenerative diseases. The study of mitochondrial genetics is complicated by the fact that isolation methods must be carefully considered to avoid inadvertent *in vitro* DNA damage. Likewise the presence of nuclear mitochondrial pseudogenes can make PCR based methods tedious and can introduce confounding into study data. Various approaches that have been developed for isolating and studying mitochondrial DNA are evaluated and modified in this thesis in order to explore two aspects of mitochondrial genetics that may

influence disease (mitochondrial DNA damage and mitochondrial mutational burden).

As outlined in figure 1.9 AD is a complex disease that involves the interaction of multiple cellular pathways many of which have been shown to modify risk for disease. Identification of novel risk factors and evaluation of known risk factors is important to improve the understanding of the disease in order to develop therapeutic interventions.

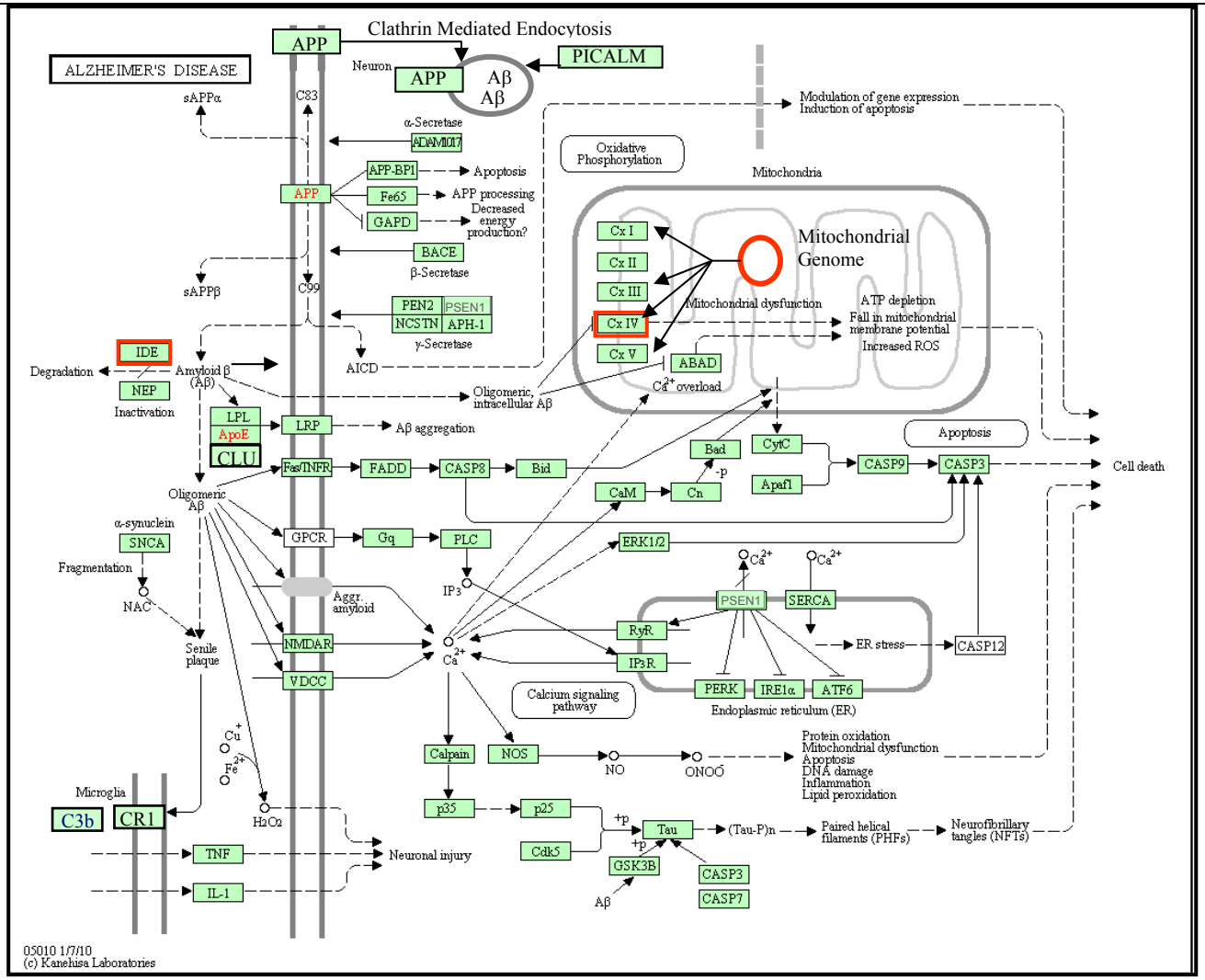

Fig 1.9 Alzheimer's Disease KEGG pathway figure. Figure adapted from KEGG website [313-315]. Areas of Alzheimer's disease pathophysiology studied in this thesis are highlighted in red.

Chapter 2.

2.1 Background

This chapter describes a study that makes use of isolated populations for both performing a genome-wide linkage study and SNP based GWAS of an established LOAD endophenotype, plasma Aβ levels. The study design combines three established approaches to the study of complex genetic diseases: (i) the use of isolated populations (see Section 1.12.2.); (ii) the use of a disease endophenotype (or intermediate phenotype (see Section 1.12.1); and (iii) the use of genome-wide linkage and association studies, similar to those discussed in Section 1.11.2.

The aim of this chapter is to identify quantitative trait loci influencing plasma Aβ concentrations that provide mechanistic insights into the regulation of plasma Aβ, candidate loci for association studies in LOAD and new therapeutic targets for future pharmaceutical development.

The populations in this study are sampled from the islands of Vis and Korcula, which are part of a collection of islands off the coast of Croatia in the Adriatic Sea, collectively known as the Dalmatian islands (Figure 2.1). Researchers at the University of Split, University of Zagreb, Institute for Anthropological Research in Zagreb, MRC Human Genetics Unit and University of Edinburgh have been studying the populations from the Dalmatian islands for some time and have found these populations to exhibit clear evidence of being isolated, both from one another and from large, outbred populations [265, 316, 317]. This collaborative effort has been referred to as the "10,001 Dalmatians study" [266] which involves the

collection of blood and data from subjects on the islands of Dalmatia. The collection and study of individuals from the islands of Vis and Korcula is part of this larger study. The subjects participating in the Vis and Korcula studies provided blood samples for biochemical measurements, provided answers to interview questions regarding, for example, smoking and dietary habits, and also allowed measurement of anthropometric traits, such as height and weight. Altogether data for over 300 traits was gathered from the study subjects and blood and plasma was aliquoted for additional studies such as this one. The subjects from the island of Vis were initially genotyped for 810 microsatellite markers to facilitate genome-wide linkage analysis for the traits of interest. The subjects from both Vis and Korcula were then also genotyped for over 300,000 SNPs to facilitate genome-wide association studies (GWAS), with a much increased marker density, for association with the traits that were collected or measured.

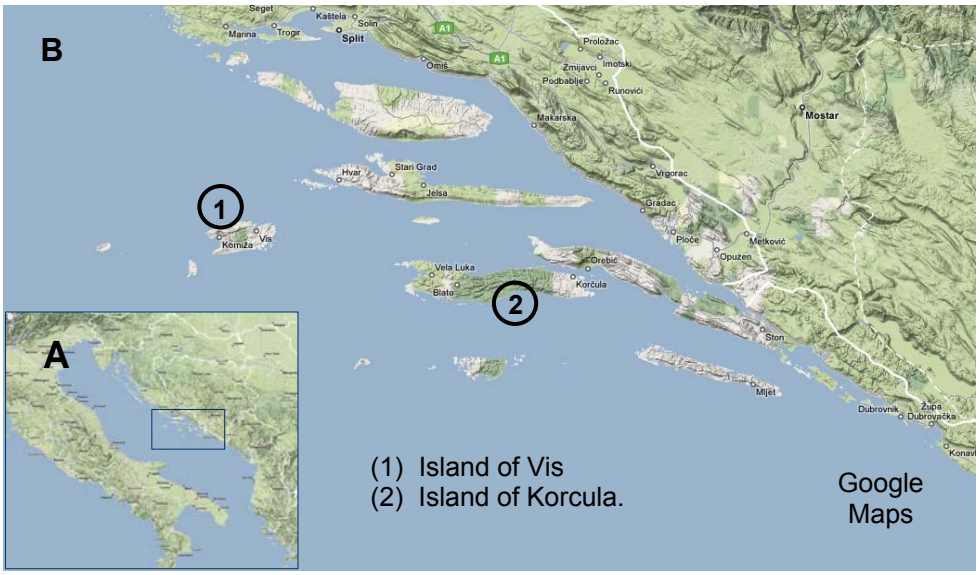

Figure 2.1. A. The location of the Dalmatian islands in the Adriatic sea. **B.** The islands of Vis and Korcula are highlighted in the collection of Dalmatian Islands off the coast of Croatia

To measure plasma Aβ concentrations, we used a well established sandwich enzyme-linked immunosorbent assay (ELISA) to specifically measure Aβ40 and Aβ42 [94]. The capture/detection antibody pairs used in the ELISA systems varied slightly from those used by Suzuki et al [94] and were different in the two studies. Details regarding the antibody pairs employed in each study can be found in section 2.2.5 and are illustrated in Figure 2.2. In both the Vis and Korcula samples the BNT77/BC05 capture/detection antibodies were used to detect Aβ42. The BNT77 capture antibody is always employed by the Younkin laboratory to assay plasma Aβ42. This is because the BAN50 capture antibody, when paired with the BCO5 detection antibody, suffers from non-specific cross-linking by proteins present in some plasma samples, which produces a non-specific signal referred to as interference. Although it is less problematic, interference can also occur using the BAN50/BA27 pair which was used to measure plasma Aβ40 in Vis [Linda Younkin, unpublished data]. For this reason the BNT77 capture antibody paired with the BA27 detection antibody was used to measure plasma Aβ40 in Korcula samples. Additional work conducted by Linda Younkin [data unpublished] indicates that the highest extreme values in most plasma Aβ studies, despite the use of antibody pairs that minimise this phenomenon, are frequently the result of interference. This appears to be unique to a small subset of subjects and does not represent genuinely high Aβ values. For this reason, criteria were devised (See Section 2.2.6) for defining the extreme values for Aβ40 and Aβ42 in both the Vis and Korcula

populations and subsequently a small number of subjects were removed from each study that had values in the highest extreme for their sample population.

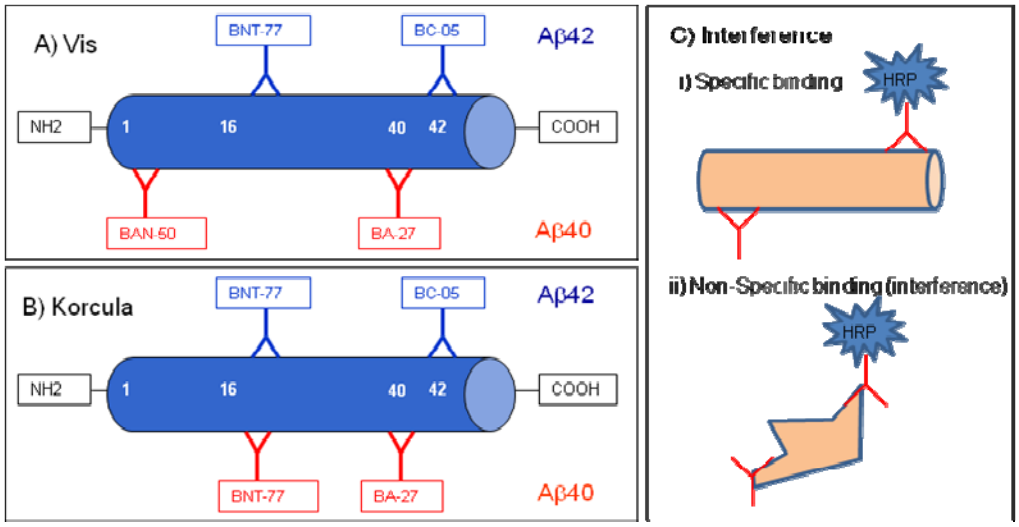

Figure 2.2 Measurement of plasma Aß A: Antibody pairs for measuring plasma Aβ40 and 42 in subjects from the Vis study. **B**: Antibody pairs for measuring plasma Aβ40 and 42 in subjects from the Korcula study **C.** Interference: i) Specific binding is shown where the variable chains specifically bind to the target peptide via the Fab region. ii) Non-specific binding is illustrated where cross-linking substances present in plasma bind to regions of the antibody not usually involved in target peptide binding.

2.2 Methods

2.2.1. Study Samples

Both population studies (Vis and Korcula) received ethical approval from Croatian and Scottish ethics boards and all participants gave informed consent. Subjects were unselected volunteers, aged 18 years and over, who were invited to participate in the studies through various media outlets including mail, posters and radio advertisement. For the Vis study, subjects were recruited from two towns on the island; Vis and Komiza [272]. Subjects participating in the Korcula study were recruited from the town of Korcula and additional villages on the eastern side of the island [270]. Subject recruitment was carried out from the year 2002 onwards by employees of the School of Public Health at the University of Zagreb Medical School, the Institute for Anthropological Research in Zagreb, Croatia, Croatian Centre for Global Health, Faculty of Medicine, University of Split, Croatia, the MRC Human Genetics Unit, Edinburgh, UK, and the Centre for Population Health Sciences, University of Edinburgh, UK.

2.2.2. Sample Collection

Fasting blood samples were collected between 8am and 9am from the Vis subjects and between 8.30am and 9.30am from the Korcula subjects. Serum and plasma were separated, aliquoted and stored at -80^{o}C until thawed for analysis or DNA isolation. The Vis plasma samples were stored as EDTA-anticoagulated aliquots and thawed only once prior to plasma Aβ measurement. The Korcula plasma samples were stored as 2ml EDTA aliquots and thawed once prior to the initial plasma Aβ

measurement. Repeat measurements of plasma Aβ, when necessary, were subjected to an additional thaw cycle in the Korcula subjects.

In addition to providing fasting blood samples; subjects also completed questionnaire interviews and underwent physical examinations in order to collect data on health, lifestyle, psychological and anthropometric measures such as BMI and height. Sample and anthropometric data collection were carried out from 2003 onwards by qualified individuals employed by the School of Public Health of the University of Zagreb Medical School, the Institute for Anthropological Research in Zagreb, Croatia and the MRC Human Genetics Unit, Edinburgh. I was not involved in the sample collection.

2.2.3. Plasma sample transfer

Plasma samples collected from Vis and Korcula study subjects were transferred to the MRC Human Genetics Unit in Edinburgh, where they are stored at -80^{o}C. A Material Transfer Agreement (MTA) was set-up to facilitate transfer of aliquots of plasma from the MRC Unit in Edinburgh to the Mayo Clinic in Jacksonville, Florida, where measurement of plasma Aβ levels was carried out. Plasma samples for the Vis study subjects (3x500µl aliquots per sample) were transferred in April 2006, Korcula plasma samples (1x 2ml aliquot) were transferred in September 2008. Samples were shipped on dry ice via Federal Express and the samples all arrived intact and still frozen. Samples were immediately placed in a freezer at -80^{o}C, where they were stored until thawing for measurement.

2.2.4. Genotyping

DNA isolation: DNA was isolated, by others, from blood samples collected from study participants using Nucleon DNA Purification kits (Tepnel). DNA was stored at -20^{o}C at the MRC Human Genetics Unit in Edinburgh.

Microsatellite markers: The Vis study samples were genotyped using an ABI Prism Linkage Mapping set for 810 microsatellite markers that define a ~5 centiMorgan (cM) resolution map (HD-5). The genotyping panel(s) and reagents were purchased from Applied Biosystems and the genotyping was carried out, by others, using an ABI Prism 3100 at the MRC Human Genetics Unit. Following implementation of quality control of the genotyped markers (by others), 746 markers were included in the analysis.

SNP genotyping: In the Vis study 991 samples were genotyped for over 317,000 SNPs using the HumanHap300-Duo Genotyping BeadChip (Illumina), of these plasma was available for 965 samples. In the Korcula study 953 samples were genotyped for over 370,000 markers using the HumanCNV370-Duo DNA Analysis Bead Chip (Illumina), although plasma was only available for 941. Genotyping of Vis and Korcula samples on Illumina platforms was carried out by the Wellcome Trust Clinical Research Facility at the Western General Hospital. Genotypes were scored, by others, using the BeadStudio Software v.3 (Illumina). Additional SNPs of interest were genotyped, by others, using Applied Biosystems Taqman assays (Applied Biosystems 7900 Real-Time PCR System) at the MRC Human Genetics Unit.

2.2.5. Plasma Aβ measurements

Plasma Aβ measurements were made using a sandwich ELISA system that has been described previously [94]. Prior to assaying the plasma, ELISA plates, buffers, antibodies and synthetic and biological standards were prepared as described below.

Antibodies: All antibodies were provided by Takeda Pharmaceutical Company Ltd. Two capture antibodies (BAN50 and BNT77) and two detection antibodies (BA27 and BC05) were used over the course of these experiments, details of which can be found in Table 2.1.

Name	Type	Source	Isotype	Purpose	Recognize
BAN50	Mouse monoclonal	Takeda	IgG1k	Capture	Aβ 1-16
BNT77	Mouse monoclonal	Takeda	IgAk	Capture	Aβ 11-28
BA27	Mouse monoclonal	Takeda	IgG2ak	Detection	Aβ 40-COOH
BC05	Mouse monoclonal	Takeda	IgG1k	Detection	Aβ 42

Table 2.1. **Antibodies used in the measurement of plasma Aβ levels in the populations samples from Vis and Korcula.**

The detection antibodies (BA27, BC05) were HRP-conjugated using a kit from Pierce (Cat#31489), making use of D-Tube Dialyzers from Novagen (Cat# 71507-3). For both preparations, the conjugate was dialyzed overnight in 1xPBS at 4^{o}C. All detection antibodies were stored at 4^{o}C and used within one month, following HRP conjugation. The capture antibodies were diluted in Sodium Carbonate Coating buffer (0.1M $NaHCO_3$, 0.1M Na_2CO_3, pH 9.6) at a concentration of 5ug/ml. ELISA plates were provided by Thermo (Immulon 4 HBX 96 well, flat bottom microtiter plates with extra high binding, Cat# 3855). 100µl of the diluted capture antibody

was immobilized on the centre wells of the plate (60 wells). The plate was then covered with an adhesive plate seal and incubated overnight at 4°C. The following day, the capture antibody was carefully removed from each of the wells and each well was washed twice with 1X PBS using a plate washer (Nunc-Immuno™ Wash 8). 300µl of Block Ace solution was then added to each of the centre wells, the plate was covered with an adhesive seal and stored at 4°C. All plates were used within 1 month.

Assay	Capture Antibody		HRP-detection antibody		Species Detected
	BAN50	BNT77	BA27	BC05	
Vis Aβ40	X	-	X	-	Aβ(1-40)
Vis Aβ42	-	X	-	X	Aβ(x-42)
Korcula Aβ40	-	X	X	-	Aβ(x-40)
Korcula Aβ42	-	X	-	X	Aβ(x-42)

Table 2.2. Pairs of antibodies used in the sandwich ELISAs and the specific Aβ species detected.

To measure plasma Aβ40, different pairs of antibodies were used for Vis and Korcula, as shown in Table 2.2. The reason for using different pairs of antibodies (interference) has been explained in section 2.1. The detection antibody BA27 specifically detects Aβ species ending at Aβ40 whereas BC05 detects Aβ species ending at Aβ42. BAN50 recognizes an epitope near the amino terminus of Aβ, so the BAN50/BA27 sandwich used for Vis samples detects primarily Aβ1-40. BNT77 recognizes Aβ11-28 so BNT77 captures Aβ species that are amino-terminally truncated (e. g. Aβ3-40 and Aβ3-42) in addition to full length Aβ species. For this

reason, the Aβ detected by the BNT77/BA27 pair and the BNT77/BC05 pair is designated Aβx-40 and Aβx-42 respectively.

Buffers: Several buffers are used during the sandwich ELISA protocol; these were all prepared prior to experimentation.

- Coating Buffer: 0.1M $NaHCO_3$, 0.1M Na_2CO_3 [pH9.6]
- EC Buffer: 0.02M sodium phosphate buffer [pH 7.0], 2mM EDTA, 0.4M NaCl, 0.2% BSA, 0.05% CHAPS, 0.4% BlockAce (Serotec), 0.05% NaN_3, pH 7.0.
- Buffer C: 0.02M sodium phosphate buffer [pH 7.0], 2mM EDTA, 0.4M NaCl, 1% BSA, 0.002% thimerosal, pH 7.0.
- 10X PBS: KCl 2g, KH_2PO_4 2g, Na_2HPO_4 11.4g, NaCl 81.2g, pH7.4, dilute in 1L deionised water.
- PBS + Tween: 0.5ml 0.05% Tween in 1000ml 1X PBS.
- Block Ace: 1% Blockace (4 g = 1 packet) in 400 ml 1X PBS, 0.05% Na azide, pH 7.4

Plate design: The established plate design for measuring plasma Aβ levels involves the use of only the middle 60 wells on the plate for measurement as a fluctuation in readings had previously been observed in the outer wells, probably due to spurious evaporation. All experimental samples and standards are measured in duplicate. This allows for 20 experimental samples, a standard curve generated from 5 known concentrations of synthetic standards and a negative standard, as well as four internal plasma standards to be measured on each plate. Each assay (Aβ40 or Aβ42)

is measured on one plate and so two ELISA plates are required to measure both peptides for every 20 experimental samples.

Synthetic standards (standard curve): Synthetic Aβ peptide for use in the standard curve was prepared, by Linda Younkin, from stocks received from rPeptide (Recombinant), Aβ1-42 Cat# A-1002-1, Aβ1-40 Cat# A1001-1. The stocks were diluted into EC buffer according to the directions from rPeptide and stored as 300µl aliquots at a concentration of 5pmol/ml. For the Vis study a standard curve range of 800 fmol/ml to 3.125 fmol/ml was implemented, using a serial dilution of 0.25 per standard (800, 200, 50, 12.5 and 3.125 and 0 fmol/ml). The 800 fmol/ml standard is at the top end of the dynamic range of the assay and very few samples fall between 800 fmol/ and 200 fmol/ml, so in order to increase the specificity of the curve the 800 fmol/ml standard for the Korcula study was eliminated and replaced with the following set of values for the standard curve for Korcula - 200, 50, 12.5, 3.125, 0.6125 and 0 fmol/ml - where a 0.25 serial dilution was likewise adopted. In both studies, EC buffer was used as the diluent.

Plasma Standards: To correct for plate to plate variation, four plasma samples prepared from four anonymous donors prior to each study were measured in duplicate on every plate. All experimental values on each plate were normalized to the mean value of the four standards on that plate. The aliquots of plasma used for a single study are all from a single blood draw but may differ in draw date. Three plasma standards were consistent between the Vis and Korcula studies and one subject differed between the two studies; this does not impact the interpretation of the results because there is no cross-study normalization.

Sandwich ELISA protocol: The prepared ELISA plates coated with the relevant capture antibody were washed twice with 1X PBS, using a plate washer (Nunc-Immuno^{TM} Wash 8). Immediately following the wash step, 50µl of EC buffer was added to the centre 60 wells to prevent the wells from drying. 100µl of the relevant standard was added to each of the standard curve wells and the plasma was diluted 1:3 directly into the wells on the plate (33µl plasma, 66µl EC). Following incubation overnight at 4^{o}C, the wells were washed twice with 1X PBS and then 100µl of prepared HRP-conjugated detection antibody, diluted in buffer C, was added to each centre well, followed by a four hour incubation at room temperature. Developer solution, at room temperature, was prepared by mixing 1:1 TMB peroxidise substrate solution and peroxidase solution at room temperature. Following incubation with secondary antibody, the plates were washed twice with 1X PBS and then washed a third time with 1X PBS + 0.05% Tween. 100µl developing solution was added to all 96 wells on the plate and Aβ40 plates were left to develop for 30 minutes whilst Aβ42 plates were allowed to develop for 45 mins. The developing reaction was terminated by the addition of 100µl 1M phosphoric acid (H_3PO_4), which raises the OD values approximately 3 fold. The plates were then read at 450nm using the Molecular Devices *Emax* precision microplate reader.

Plate reader: The plate design, standard curve concentrations and dilution of experimental samples were entered into the Softmax software package, which facilitates calculation of plasma Aβ values (fmol/ml) for the experimental samples and internal plasma controls on each plate. The data was saved and exported from this software into text format.

Plate normalization: The mean value of the four plasma standards was calculated for each plate and the values for the experimental samples on each plate were normalized against the average of these mean values in the study. For example, if plate A has a mean value for the internal controls of 100 fmol/ml, and the mean for all the plates in the study was 120 fmol/ml, then the experimental values on plate A would be increased by 20% to fall in line with the overall study values.

Quality control: Various quality control measures were put in place in order to prevent experimental variation or error from unduly influencing the results. As mentioned previously, four internal control plasma samples were included on each plate to facilitate normalization of each plate due to plate-to-plate variation. When a plate failed to successfully measure all four standards in duplicate, the whole plate was discarded and repeated. When more than one synthetic standard failed to have duplicate measures in the standard curve, the whole plate was again discarded and repeated. Finally, when the duplicate measurements of the experimental plasma samples had a coefficient of variation (CV) of greater than 0.2, these were also repeated. When repeats due to a high CV were necessary and more than 2 measurements were recorded for an experimental sample, all measurements were used but in order to avoid the undue influence of extreme high or low values the median rather than the mean of all values was used.

2.2.6. Plasma Aβ analysis

Descriptive analysis of raw values: The plasma values for Aβ40 and Aβ42 were used to create a total of 4 plasma Aβ traits for further analysis, these are Aβ40,

Aβ42, the ratio of Aβ42:Aβ40 (from here on out referred to as "ratio") and the sum of Aβ40+Aβ42 (from here on out referred to as "total"). The mean, median, standard deviation and distribution of values for each trait are described in the results (Table 2.3.). Due to interference, in which falsely extreme high values can occur, the highest extreme values for both Aβ40 and Aβ42 were removed. The highest extreme values for plasma Aβ40 and 42 were determined as those with an absolute value of greater than 3 times the standard deviation (SD) above the mean for all measured values in the population: The values for all four traits were rank transformed as the distributions did not fit a Gaussian curve. The rntransform function available in the R package GenABEL [318] was used to rank transform each of the traits.

Correlation of plasma Aβ traits: Using Microsoft Excel, Pearson's test for correlation was used to test for the level of correlation between the plasma Aβ40 and Aβ42 traits in each of the two populations.

2.2.7. Linkage analysis using SOLAR

Pedigrees were established from family relationships reported during interviews/questionnaires with the subjects. Multipoint identity-by-descent (MIBD) estimates were generated by Dr. Veronique Vitart using Markov Chain Monte Carlo (MCMC) methods implemented in the program Loki [319, 320]. The files generated were used for subsequent polygenic and multipoint linkage analyses in SOLAR [321], of Aβ40, Aβ42, the Aβ42/Aβ40 ratio, and the Aβ40+Aβ42 sum. The covariates of age, sex, age*sex, age^2, age^2*sex, and *APOE* ε4 alleles

(presence/absence) were investigated for association with each of the four plasma Aβ traits, using the 'polygenic –screen' function in SOLAR. Those variables with a p-value of less than 0.1 were retained as covariates in the models analyzed by SOLAR. The polygenic analysis provides an estimate of the heritability of the trait of interest (H2) and analyzes each covariate for association with the trait. The polygenic analysis also creates a null0 model which is used in subsequent linkage analyses, where the null hypothesis tested is that the additive genetic variance due to the QTL is equal to zero; i.e. no linkage. The null hypothesis is tested using variance components linkage analysis where the LOD score represents the difference between the $\log_{10}$ likelihood of this null model and a model where the variance due to the QTL is estimated. The null hypothesis can be accepted or rejected based on these LOD scores, where the well established threshold of a LOD score greater than 3.3 traditionally reflects rejection of the null hypothesis with a p-value of <0.05. Multipoint linkage analysis was run for all 22 autosomal chromosomes implemented in SOLAR. Results were plotted using Microsoft excel and loci with a LOD score greater than 1 were mapped based on Haldane's mapping function. The UCSC genome browser was used to find the base pair positions of the closest markers to marker loci with a maximum LOD >1 and likewise any genes in the proximity (within 0.5 LOD support interval or ±10cM, whichever is smaller). The LOAD linkage results from the recently published Butler *et al.* paper [2] were evaluated for overlap with the linkage results from this study and are summarized in Table 2.5.

2.2.8. Genome-wide association analysis: GenABEL

Generation of objects gwaa.data-class: The R package GenABEL [318] and its functionalities were used for genome-wide association analysis of each of the four plasma Aβ traits in each population. Genotype data, following quality control carried out using the Illumina Beadstudio software, was formatted into GenABEL internal format, by others, using the functionalities within GenABEL that facilitate conversion of genotype, SNP and subject data into GenABEL format, using the various "convert" functions. Using the converted genotype data provided as a '.raw' file and phenotype data (plasma Aβ), a QC script was executed in R. This script facilitated creation of the object class gwaa.data.class for the population studied (Vis or Korcula) through incorporation of the phenotype and genotype/subject data. This script also generates a 'genomic kinship' or 'gkin' matrix, based upon identity-by-descent (IBD) estimates using autosomal SNP markers, which serves as a replacement for the traditional reported pedigree method. Finally, this script also incorporated quality control measures, such that the object created included only data that met our quality control criteria.

Quality Control (QC): A sequential QC procedure was executed for each population. SNPs were removed which had a minor allele frequency of less than 2%, conservative sample and marker call rate limits of 97% and 95% respectively were implemented and a lenient Hardy-Weinburg p-value threshold of 1E-10 was used. To avoid inadvertent sample duplication and to control for the presence of any twins, we removed subjects who had IBS greater than 95%. The quality control sequence was run successively three times to take into account changes in the above

statistics following removal of SNPs or subjects. Population outliers were also identified and removed when exceeding certain limits using multidimensional scaling (mds) methods also implemented in the QC script.

Trait analysis: As mentioned in Section 2.2.6, the plasma Aβ measurements for each of the populations did not fit a Gaussian distribution and so the 'rntransform' function available in the R package GenABEL was utilized to rank transform the data distributions. Heritability estimates and the association of covariates (age, sex and *APOE* ε4 allele) with each of the traits were evaluated using the polygenic function in GenABEL. Genome-wide association analysis, using an additive model, was carried out using the mmscore function in GenABEL, including age and sex in the model regardless of significance.

Polygenic analysis: The polygenic model takes on the form shown below. As an example of this model, analysis of the plasma Aβ40 trait in Vis, with the covariates of age and sex is shown:-

```
AB40 <- polygenic(rnkqt40~age+sex, kin=gkin, df1)
```

This creates an object 'AB40' which is an estimation of the polygenic model for the rank transformed plasma Aβ40 trait (rnkqt40) conditional upon the covariates age and sex and incorporates the genomic kinship matrix information created previously (kin=gkin).

The maximum likelihood estimates (MLEs) for the (narrow-sense or additive) heritability and significance of each covariate are included within the results of the

polygenic analysis. The significance of each covariate was evaluated by running the model with and without the covariate included in the model. Likewise, to determine the significance of the heritability of the trait, the full final model with heritability set to 0 (h^2=0) was also run. A one-sided p-value can be obtained for the heritability estimates since a negative heritability is impossible. To assess population stratification, the polygenic model was also used to estimate the genomic control parameter, λ.

GWAS analysis, mmscore: For future meta-analysis purposes the mmscore GWAS analysis was incorporated into the 'formetascore' function. For example:-

```
metatraitVis40 <- formetascore(AB40, df1, stat=mmscore, transform="no",
build=36, verbosity =2)
```

The above command will run mmscore GWAS analysis on the AB40 object created earlier under the polygenic function, whilst also setting up the output file for future meta-analyses within MetABEL [318].

MetABEL: Meta-analysis of the two populations was executed using the functionalities available in the associated R package MetABEL. To perform meta-analysis of the Vis and Korcula Aβ40 GWAS data, which was stored as two data frames named "metatraitVis40" and "metatraitKor40", the following script was used:-

```
Meta40 <- metagwa.tables(metatraitVis40, metatraitKor40, name.x="Vis",
name.y="Kor")
```

The resultant data frame containing the meta-analysis results was then written to a csv file.

Quantile-quantile (QQ) plots: QQ plots were generated, representing the distribution of observed association p-values from each GWAS locus, against the distribution of expected p-values under the null hypothesis of no association, using plotting functions in R.

Haploview: The results of the GWAS were viewed using Haploview 4.1 [322] by importing the GWAS results in "PLINK" format. Haploview recognizes specific column headings in the PLINK format that do not match the GenABEL output, so in order to view the data a Perl script was used to manipulate the data so that it was in a format suitable for uploading into Haploview. The substitute command in Perl was used to alter the csv file from a comma delimited to a tab delimited file, and to substitute the column header 'name' for 'SNP' and the column header 'chromosome' for 'CHR'. The file was converted from .csv to .txt and was then ready for import into Haploview. The plot function in Haploview was used to generate Manhattan plots of the results for each trait in each population and for the meta-analysis.

2.2.9 Convergent analysis of LOAD and plasma Aβ GWAS

The SNP's identified in this chapter, with a p-value of 1E-05 for association with one of the plasma Aβ traits (Tables 2.7, 2.9 and 2.10), were evaluated for association with LOAD. Almost all of these SNPs had previously been genotyped in a LOAD GWAS [240] and this data was queried for 19 of the 20 SNP's listed in tables 2.7, 2.9 and 2.10. One SNP rs6939603 was not present in the LOAD GWAS queried. Analysis was implemented in PLINK as described by Carrasquillo et al.

Conversely, SNPs that have been identified in recent GWAS [241, 242] proximal to the genes CLU, CR1 and PICALM, were evaluated for association with plasma Aβ levels in the populations described in this chapter by likewise querying the results generated from the GWAS analysis.

2.3 Results

2.3.1 Plasma Aβ measurements

Vis: Plasma Aβ40 was successfully measured for 999 subjects. The analysis of two full plates was repeated due to unreliable or outlier internal standard values. 205 subjects had a CV greater than 0.2 in the initial measurement and so were repeated in duplicate and the median value calculated from 3 or 4 measurements. For the plasma Aβ42 assays, 998 subjects had successful measurements after 6 full plates were repeated, as were 350 subjects that either failed the initial measurement, or had a CV greater than 0.2 for the initial duplicate measurements. Eleven subjects had Aβ40 values greater than 3 SD above the mean (> 91.21) and 7 subjects had Aβ42 values greater than 3 SD above the mean (>62.60). There was no overlap between these two set of subjects and so 18 subjects were removed from the analysis for Vis.

Korcula: Plasma Aβ40 was successfully measured for 960 subjects. Two full plates were repeated due to unreliable or outlier internal standard values. One hundred and eleven samples had a CV greater than 0.2 in the initial measurement and so were repeated and the median value calculated from 3 or 4 measurements. For plasma Aβ42 assays, 954 subjects had successful measurements after 18 full plates and 22 subjects with a CV greater than 0.2 were repeated. For Korcula, 9 subjects had Aβ40 values greater than 3 SD above the mean (>191.16) and 8 subjects with Aβ42 values greater than 3 SD above the mean (>112.70). Of these, six were present as extreme values for both traits and so a total of 11 subjects were removed from the analysis.

The discrepancies between the number of subjects with successful measurements for Aβ40 and Aβ42, 999 subjects versus 998 respectively for Vis and 960 subjects

versus 954 respectively for Korcula, were due to the low plasma volume available for some subjects, such that there wasn't enough plasma to measure both traits multiple times.

Importantly, not all samples that were genotyped had plasma available for Aβ measurement; likewise not all samples that were measured for plasma Aβ were also genotyped. In total 949 subjects from the Vis study and 945 subjects from the Korcula study had both genotype data and plasma Aβ data qualified for GWAS analysis.

The distribution of trait values as shown in figures 2.3.1(A-D) and 2.3.2 (A-D) for each trait and both populations does not fit a Gaussian distribution. Following rank transformation, each of the traits fit a Gaussian distribution an example of which is shown in figure 2.3.3.

A	Vis Dataset	Trait	Min	Max	Mean	Median	SD	Ex Value[a]
	All	Aβ40	6.24	176.60	47.31	45.53	14.63	91.21
		Aβ42	3.03	231.71	22.14	20.00	13.48	62.60
	ExOutRem[#]	Aβ40	6.24	91.20	46.52	45.32	12.36	na
		Aβ42	3.03	59.33	21.34	19.99	8.47	na
		Aβ42:40 Ratio	0.07	2.72	0.48	0.43	0.23	na
		Aβ42 + Aβ40	17.13	127.31	67.50	65.52	16.32	na
B	**Korcula Dataset**	**Trait**	**Min**	**Max**	**Mean**	**Median**	**SD**	**Ex Value[a]**
	All	Aβ40	23.21	828.66	82.09	77.39	36.36	191.16
		Aβ42	4.77	836.23	17.17	13.22	31.84	112.70
	ExOutRem[#]	Aβ40	23.21	182.34	79.67	77.29	18.23	na
		Aβ42	4.77	110.88	15.09	13.16	8.24	na
		Aβ42:40 Ratio	0.07	1.02	0.19	0.17	0.09	na
		Aβ42 + Aβ40	43.35	285.86	94.69	91.36	21.98	na

Table 2.3. Descriptive statistics for plasma Aβ traits from the Vis (A) and Korcula (B) populations. * The trait value maximum limit, above which samples are deemed as outliers. a. Trait values with extreme outlier values were removed from the population. Units are in fmol/ml. Statistics are based upon samples that qualified for GWAS analysis.

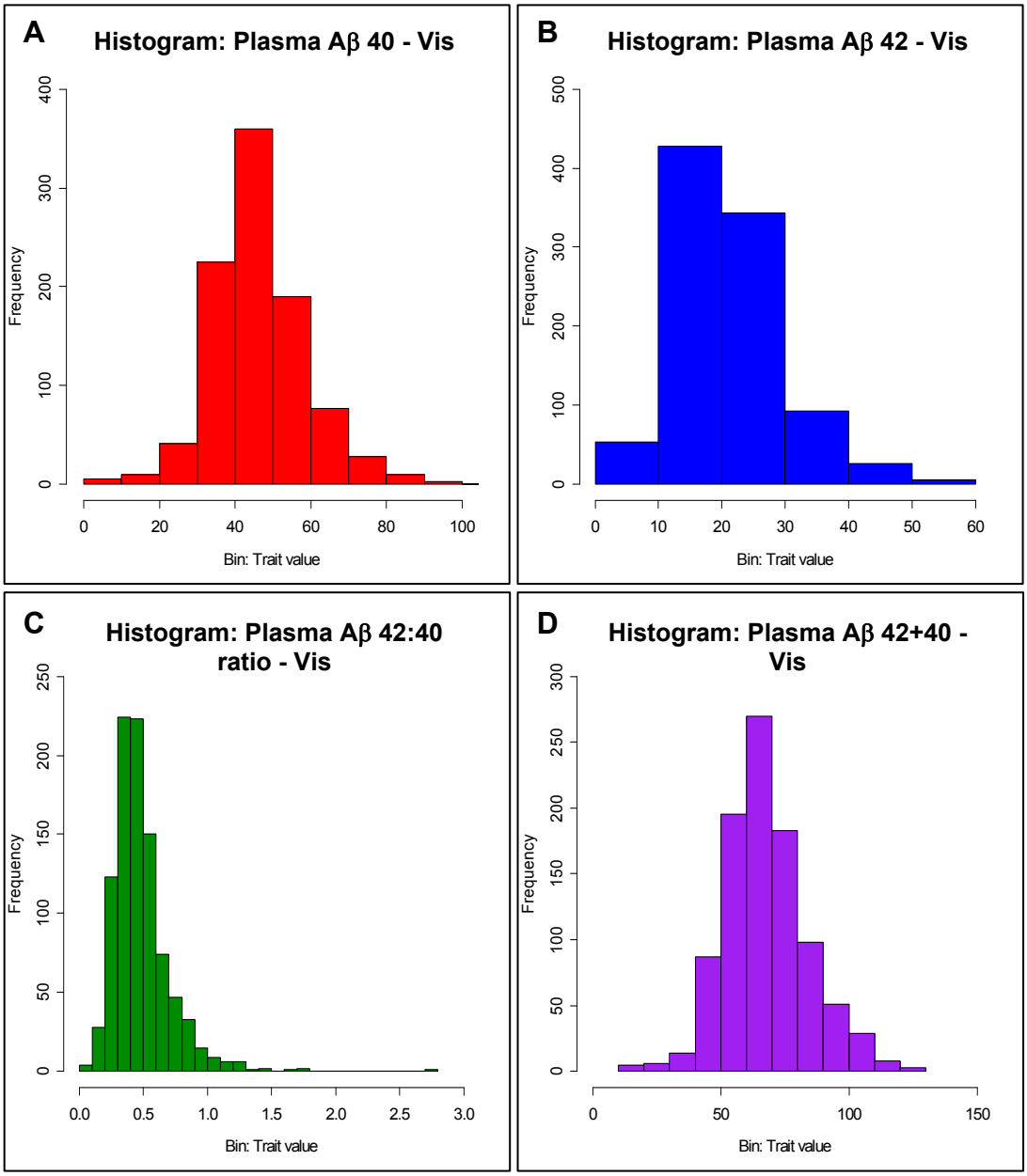

Figure 2.3.1. A-D: Histograms showing the frequency distribution of untransformed plasma Aβ traits for the Vis population. Subjects with extreme outlier values have been removed

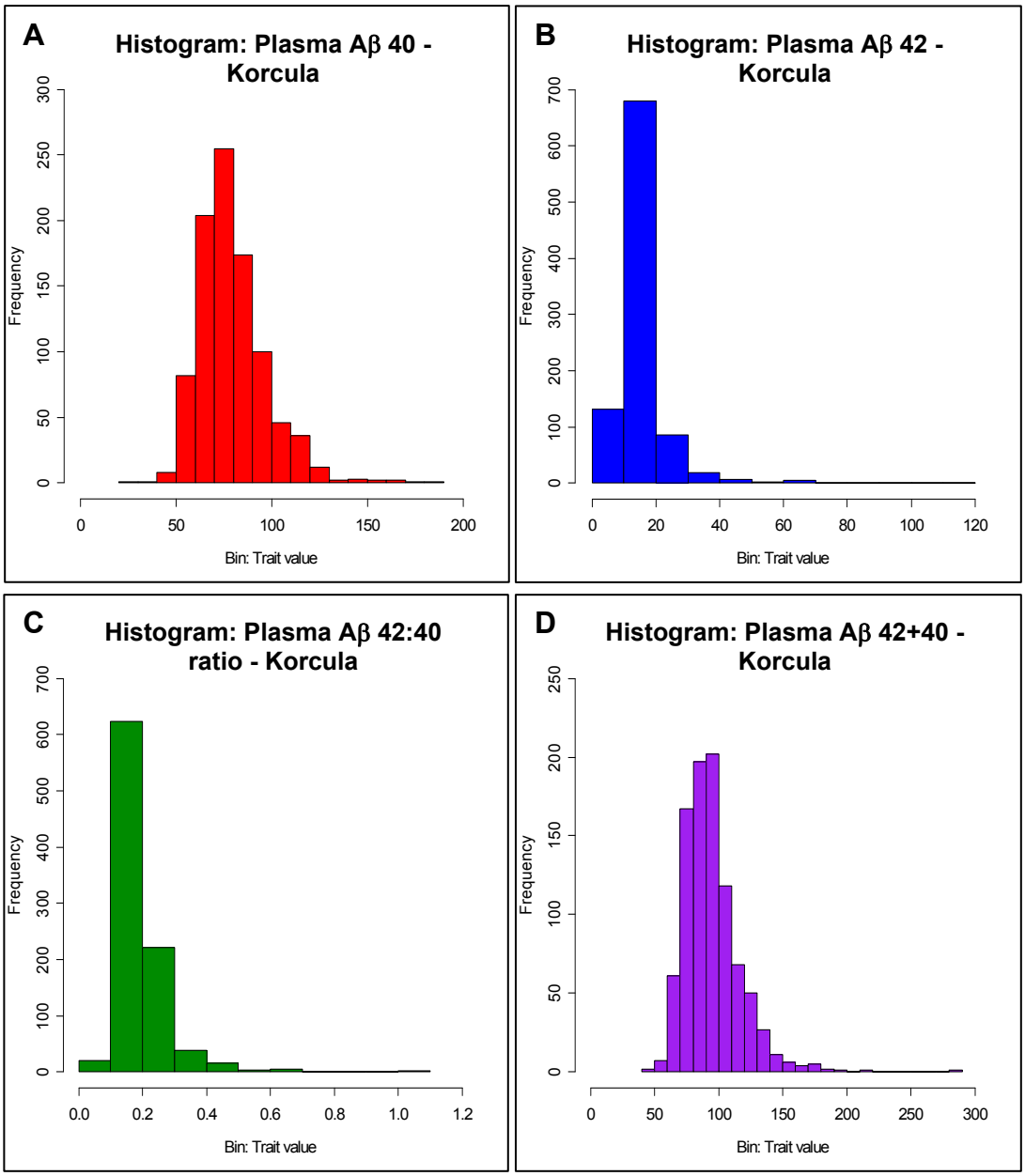

Figure 2.3.2. A-D: Histograms showing the frequency distribution of untransformed plasma Aβ traits (fmol/ml) for the Korcula population. Subjects with extreme outlier values have been removed.

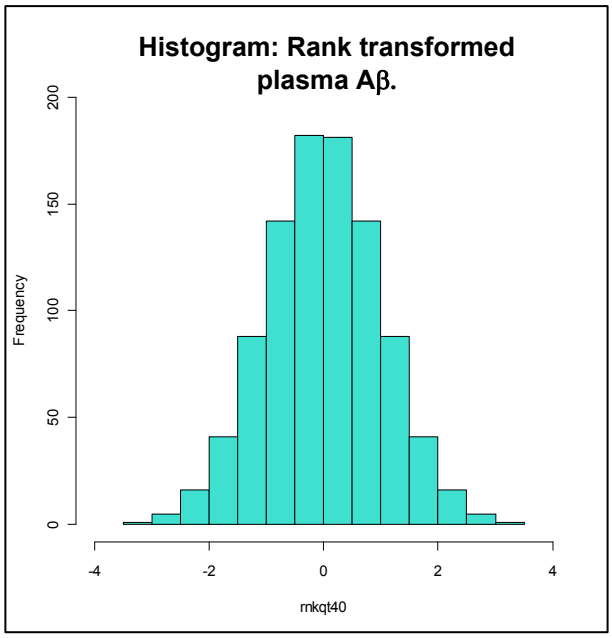

Fig 2.3.3. Histogram representing the frequency distribution of plasma Aβ traits following rank transformation. The "rntransform" function executed in the R package GenABEL was used. Each of the traits are rank transformed to a value between -3.5 and 3.5. Rank transformation to fit a Gaussian distribution is necessary to meet the assumption of normality to facilitate analysis without the need for computer intensive permutations.

Plasma Aβ trait correlation: The correlation between the two measured plasma Aβ traits (Aβ40 and Aβ42) was evaluated are using Pearson's test for correlation (r). The results indicate they are highly significantly correlated in both the Vis ($r = 0.245$, $p < 0.0001$) and Korcula ($r = 0.313$, $p < 0.0001$) populations.

2.3.2 Linkage Results in the Vis population.

Following elimination of the extreme outlier values and rank transformation of the traits, as described above, multipoint linkage analysis was performed for each of the four plasma Aβ traits, using the methodologies available in SOLAR. The polygenic analysis did not identify association with the presence or absence of an *APOE* ε4 allele for any of the traits ($p>0.22$) and so this covariate was eliminated from further analysis. The age covariate demonstrated suggestive ($p<0.1$) or significant ($p<0.05$) association with all four of the traits and so was consistently retained in each of the

models. Sex did not demonstrate any evidence of association with any of the traits and was consequently removed from further analysis based upon use of the "polygenic –screen" command. The results for the polygenic models are described in Table 2.4.

Trait	h^2 (heritability)		Significant covariates (p<0.1)	Kurtosis
	beta	p-value		
Aβ40	0.165	0.041	Age, Age^2	0.281
Aβ42	0.046	0.317	Age	0
Aβ42:40 ratio	0.127	0.078	Age, Age^2	0.124
Aβ40+42	0.114	0.132	Age, Age^2	0.088

Table 2.4. **Heritability (h^2) estimates and results for the polygenic model executed using SOLAR**. Significant covariates at a p-value <0.1 are listed and included in the final model. The residual kurtosis is also shown indicating that the trait distributions do not violate the assumption of normality.

As shown in Table 2.4, the plasma Aβ40 trait exhibited significant (narrow sense) heritability at 16.5% (p=0.041). The "ratio" and "total" Aβ traits had suggestive heritability's of 12.7% and 11.4% respectively, whilst the plasma Aβ42 trait did not exhibit any evidence of being a heritable quantitative trait. As might be expected given the low heritability's observed, none of the traits revealed a linkage peak with a LOD score greater than 3.3 which is generally considered to be the threshold for genome-wide significance. The largest peak observed was on chromosome 6 (LOD=2.34) at 150cM (Haldane's mapping function) for the plasma Aβ40 trait. The closest genotyped marker was D6S1569. In total, 10 linkage peaks with a LOD score greater than 1, which was used as the threshold for suggestive linkage, were

observed for three of the traits. All of the peaks had over 10 known genes and one or more 'AlzGene' genes within the support interval, as reported on the AlzGene website [31]. The regions of suggestive linkage are summarised in Table 2.5. There were not any regions of suggestive linkage (LOD>1) for the Aβ42 trait and this trait also exhibited the lowest heritability, $\beta = 0.046$, $p = 0.317$. The linkage results are represented in Figures 2.4.1 through 2.4.4.

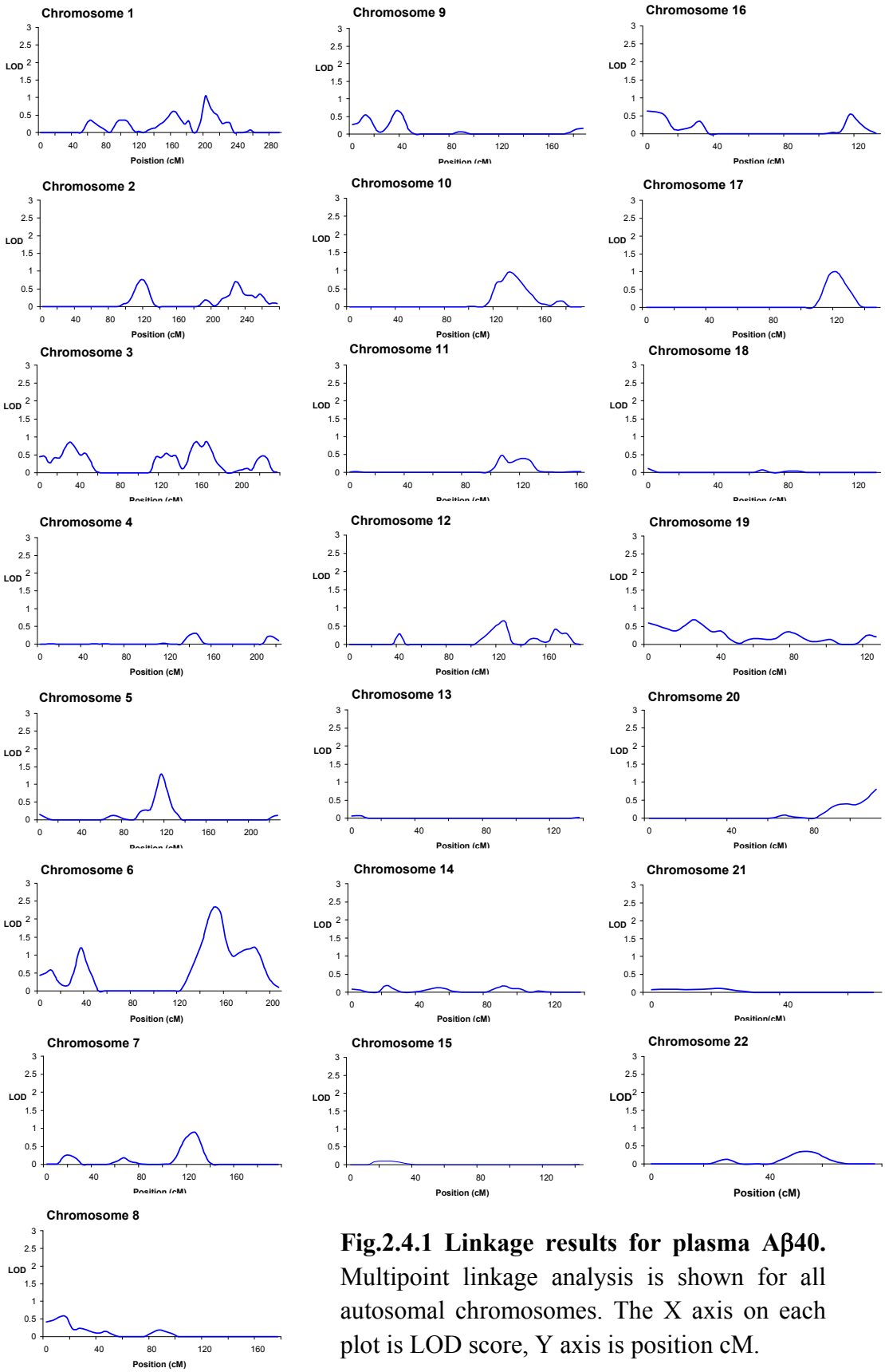

Fig.2.4.1 Linkage results for plasma Aβ40. Multipoint linkage analysis is shown for all autosomal chromosomes. The X axis on each plot is LOD score, Y axis is position cM.

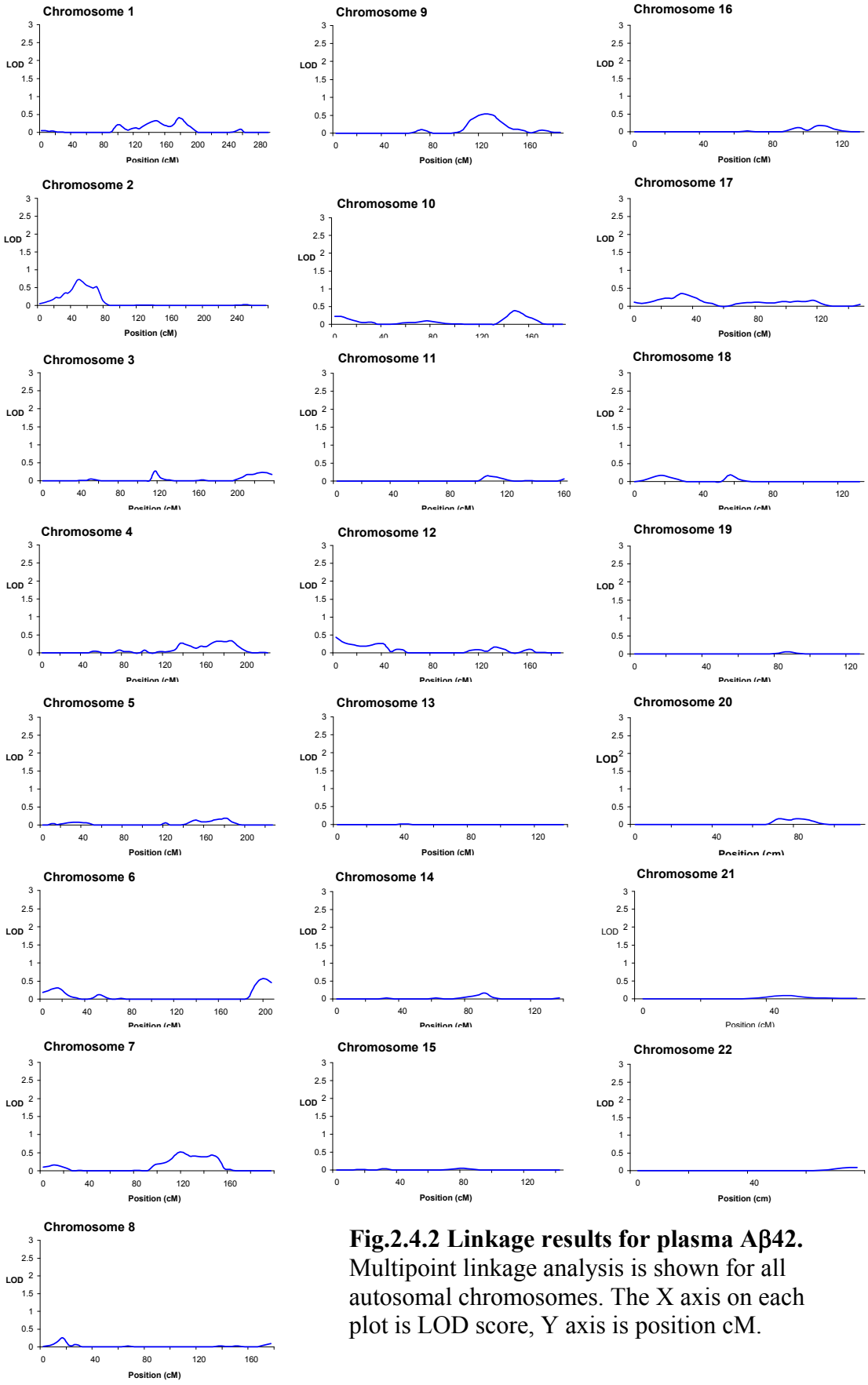

Fig.2.4.2 Linkage results for plasma Aβ42. Multipoint linkage analysis is shown for all autosomal chromosomes. The X axis on each plot is LOD score, Y axis is position cM.

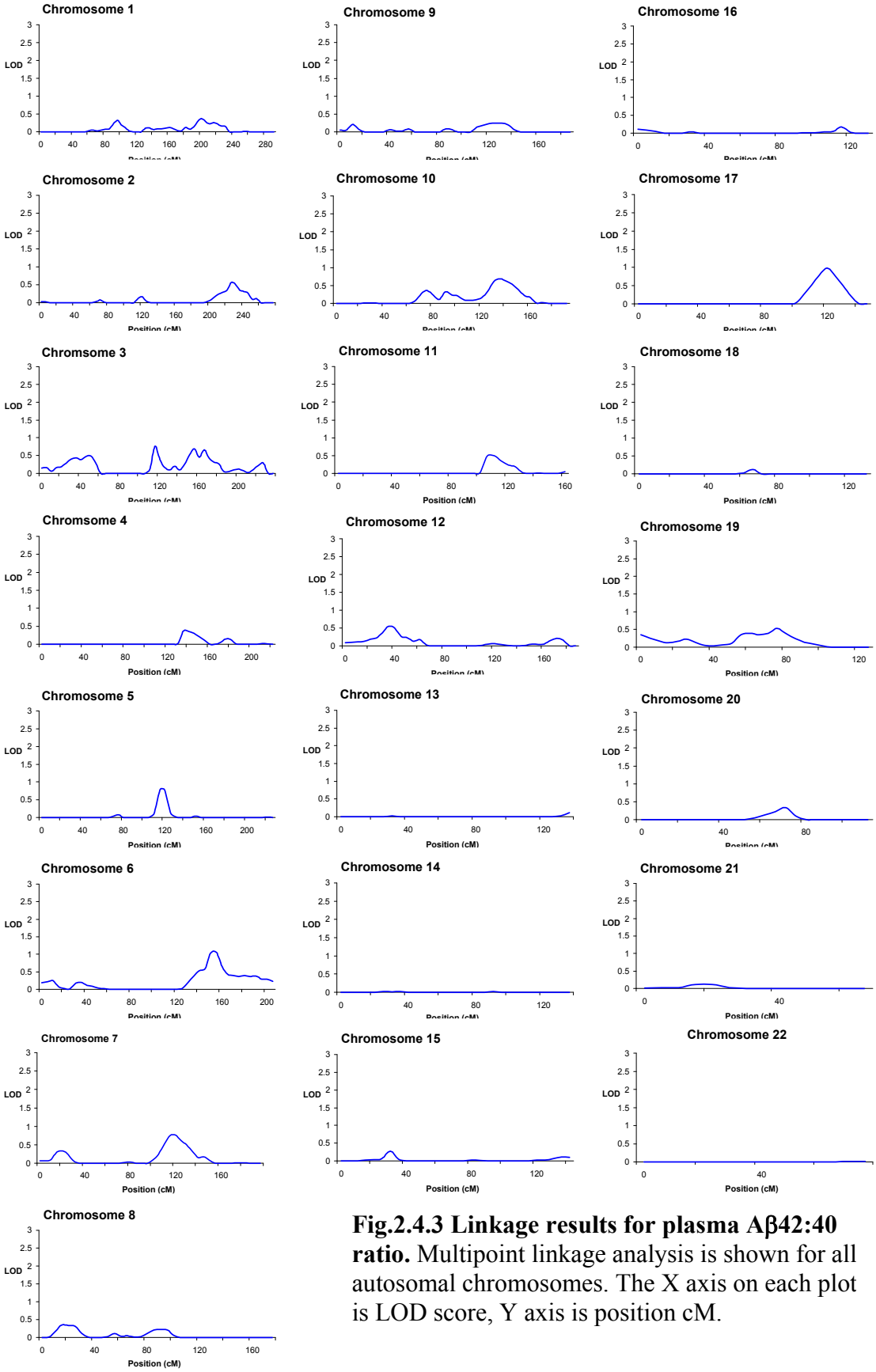

Fig.2.4.3 Linkage results for plasma Aβ42:40 ratio. Multipoint linkage analysis is shown for all autosomal chromosomes. The X axis on each plot is LOD score, Y axis is position cM.

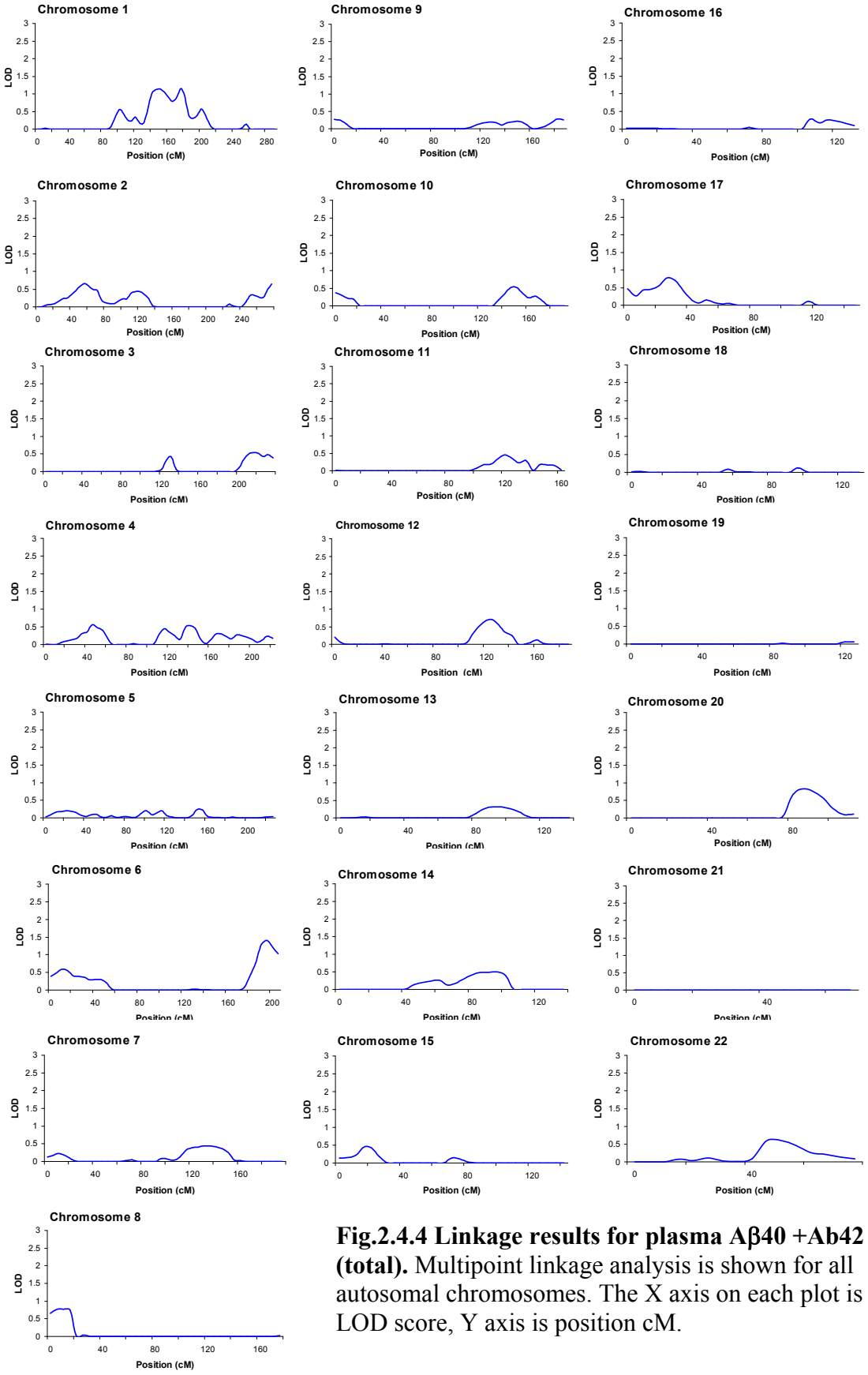

Fig.2.4.4 Linkage results for plasma Aβ40 +Ab42 (total). Multipoint linkage analysis is shown for all autosomal chromosomes. The X axis on each plot is LOD score, Y axis is position cM.

Chromosome, Band	Trait	Maximum LOD	Position (cM)	Nearest Marker	Start Position kbp	Support Interval (cM)	Support Interval (kbp)	# Genes within support interval	Alzgene genes within support interval*	Butler et al overlap?
1p13.1	Aβ42:Aβ40 Ratio	1.14	150	D1S252	117,358	140 - 160	110,985 - 157,347	>50	NGF, HMGCS2, PRKAB2, APH1A, CTSS, FAM63A, **CHRNB2**, LMNA, PMVK, FDPS, APOA1BP, GBA, NTRK1	Yes: 110.1 to 159.1 Mb
1q23.3	Aβ42:Aβ40 Ratio	1.16	175	D1S2878	163,670	165 - 185	157,347 - 172,770	>50	CRP, NCSTN, F11R, USF1, FCER1G, RGS4, APOA2, RXRG, POU2F1, PRDX6	Yes: 159.1 to 185.2 Mb
1q31.1	Aβ40	1.05	200	D1S2877	187,611	195 - 215	180,605 - 203,981	>50	CFH, PTGS2, REN	No
5q14.3	Aβ40	1.29	115	D5S618	89,802	110 - 120	85,356 - 99,658	44	CAST	No
6p23	Aβ40	1.21	35	D6S259	14,843	30 - 40	10,133 - 20,478	11	**NEDD9**, ATXN1	No
6q24.1	Aβ40	2.34	150	D6S1569	139,095	140 - 160	131,772 - 149,621	>50	ENPP1, **LOC645503**	No
6q24.1	Aβ42+ Aβ40	1.06	155	D6S308	141,298	140 - 160	131,772 - 149,621	>50	ENPP1, **LOC645503**	No
6q26	Aβ40	1.21	185	D6S305	162,115	170 - 190	153,855 - 165,989	48	SOD2, ACAT2, LPA, PLG	No
6q27	Aβ42:Aβ40 Ratio	1.41	195	D6S1697	167,764	185 - 205	162,115 - 170,394	37	RPS6KA2	No
17q25.1	Aβ40	1.00	120	D17S1807	69,782	110 - 130	65,886- 74,536	>50	GRB2	Yes: 66.6 to 78.8 Mb

Table 2.5. Linkage results summary table. Linkage loci identified with a LOD score > 1 are listed here in order of chromosome and position on each chromosome. Base pair positions are reported according to NCBI build 36 (March 2006). Linkage peaks reported by Butler *et al*. [2] are listed when there is overlap with the reported linkage peaks identified in this study. *Genes reported in the AlzGene database at the time of query (March 2010), bolded genes represent AlzGene "Top hits" genes also at the time of the query. An explanation of the AlzGene database and "Top hits" genes can be found in section 1.11.1.

AlzGene database: The AlzGene database [31] of AD association studies was used to determine if any previously studied genes had been reported within the support intervals of our linkage peaks. In total, thirteen AlzGene genes are present within the support intervals of the Aβ40 peaks and 26 AlzGene genes are present within the support interval of the Aβ ratio and Aβ total peaks. Of these genes, two are found in overlapping peaks, so we found 37 LOAD candidate genes within the support interval of the suggestive linkage loci in this study, of which three (*LOC645503, CHRNB2* and *NEDD9*) are in the AlzGene 'top genes' list (March 2010).

LOC645503 is contained within a pseudogene located in the support interval of the linkage peak on chromosome 6 with the highest LOD score in this study (Aβ40, LOD = 2.34). According to the AlzGene database (queried on March 25th 2010), rs6907175 lies within the same pseudogene and is associated with Alzheimer's disease, based upon meta-analysis of published data and ranks 21st in the Alzgene list of "top-hits" [31, 235, 323, 324]. This pseudogene is described on the NCBI website as: "similar to SUMO-1 activating enzyme subunit 2".

CHRNB2: Cholinergic receptor, nicotinic, beta 2 (neuronal) encodes the beta subunit of a neuronal nicotinic acetylcholine receptor.

NEDD9 maps to chromosome 6 and was cloned and characterized in 1996 by Law *et al*. and Minegishi *et al*. [325, 326]. It has two major isoforms and has been implicated in cancer in several studies [327-329] and in Alzheimer's disease and Parkinson's disease by genetic association studies [330-332].

Previously published linkage studies: There have been several published LOAD linkage studies [2, 333-346], as well as one study looking at plasma Aβ and linkage

to chromosome 10 [62]. In 2009 Butler et al [2] published a meta-analysis of independent LOAD linkage studies and our results have been compared with the results from this study, indicating overlapping linkage loci in Table 2.5. On chromosome 1, two peaks were identified with suggestive linkage to the plasma Aβ42:40 ratio that map to two of the top four linkage loci from the Butler paper. In addition, overlapping linkage peaks were identified on chromosome 17 indicating suggestive linkage to plasma Aβ40 that almost exactly overlaps with the Butler finding. Notably, whilst almost all published LOAD GWL (Genome Wide Linkage) studies identify linkage to the *APOE* locus on chromosome 19, none of the plasma Aß traits showed linkage to this chromosome. This is not entirely surprising given that the *APOE ε4* allele did not show significant association with any of the plasma Aβ traits in Vis. This result indicates that the *APOE* locus may have an impact on the disease that does not greatly influence circulating Aβ concentrations in this population and is perhaps independent of Aβ pathogenic pathways. Previous studies of plasma Aβ levels have likewise failed to identify significant association with the *APOE E4* genotype [204, 347].

To summarise, linkage analysis of four non-independent plasma Aβ traits measured in over 900 subjects from the island of Vis failed to identify linkage peaks with a significant LOD score of >3.3. Although no significant linkage peaks were identified, several suggestive peaks were found, the most significant of which was on chromosome 6, with a LOD score of 2.34 for plasma Aβ40. Many previously reported AD candidate genes fall within the support interval of these suggestive

peaks. Of further interest is the overlap of three of these peaks with loci reported by Butler *et al*. [2] in their analysis of published, independent LOAD linkage studies. This implies that these LOAD loci might play a role in the disease via an Aβ pathway. These loci would be of interest for further investigation into both plasma Aβ and LOAD.

2.3.3 Vis GWAS

Following the Vis linkage study, a GWAS was also conducted on these same subjects from Vis.

Quality control: As mentioned in Materials and Methods (section 2.2.8), quality control (QC) procedures were run multiple times in succession. Before QC, there were 317,503 SNPs and 991 subjects; after QC procedures, there were 308,998 SNPs and 925 subjects. The majority of the removed SNPs were due to low call rates (7,131 SNPs) with the other major reason being low minor allele frequency (1,285 SNPs). The most common reason for removal of subjects from the analysis was also low call rates (60 subjects) with a small number being removed due to high autosomal heterozygosity or IBS >95% (meaning that the subject is either duplicated in the study or has a twin).

Following QC, 93% of subjects and 91% of SNPs remaining had >99% successful genotypes. The mean heterozygosity per SNP was 0.35 and per subject was 0.34. Genetic substructure was evaluated in GenABEL, based on multidimensional scaling (MDS) and the first two principal components were plotted. Following the removal of known outliers [272], no further outliers were detected.

Polygenic model

The heritability of each trait and the trait variance associated with covariates was estimated using the polygenic function available in GenABEL. As described in Materials and Methods (section 2.2.8) the maximum likelihood estimates (MLE) for each of the covariates, the heritability's and corresponding p-values were obtained. The *APOE ε4* covariate (dichotomous variable, 0 = no *APOE ε4* alleles and 1= one or more *APOE ε4* alleles) was not nominally significant in any of the analyses ($p>0.05$) and so it was not included as a covariate in the polygenic model. Arbitrary inclusion of this covariate resulted in a reduced population sample size due to missing *APOE* genotype data in some subjects, which would lead to reduced power. Inclusion of the *APOE* data was only logical if there was evidence for association with Aβ traits. Consequently, only age and sex were included as covariates in the final polygenic models. The results for covariate and heritability estimation for each trait are shown in Table 2.6. Age was highly significantly associated with plasma Aβ40, as were the dependant traits of the Aβ ratio and total Aβ. However, despite the high level of correlation between the Aβ40 and 42 traits, there was no evidence for a significant influence of age on the plasma Aβ42 trait. Sex was not significantly associated with any of the plasma Aβ traits.

GWAS: mmscore

As described in Materials and Methods (section 2.2.8), the mmscore (FASTA) method in GenABEL was utilized to carry out the GWAS, conditional upon the estimated polygenic model. The parameter λ which is an estimation of test statistic inflation (or deflation) used for genomic control methods can also be estimated

using mmscore. For all four traits, λ was equal to 1.0 and thus the population sample shows no requirement for implementation of genomic control.

Quantile Quantile plots in Figure 2.5 (a-d) represent the distribution of the observed p-values against those that we would be expected under the null hypothesis.

Vis	Age		Sex		Heritability (h^2)	
Trait	beta	P-value	beta	P-value	beta	P-value
Plasma Aβ40	0.025	<0.0001	-0.013	0.829	0.028	0.376
Plasma Aβ42	0.003	0.221	-0.042	0.558	0.134	0.070
Plasma Aβ42:40 ratio	-0.023	<0.0001	-0.058	0.426	0.089	0.147
Sum of plasma Aβ40+42	0.019	<0.0001	-0.028	0.648	0.081	0.191

Table 2.6. Maximum Likelihood Estimate's of effect size (beta) and significance (p-value) in the Vis population for covariates and heritability of Aβ quantitative traits. The results were estimated using the polygenic function in GenABEL. The beta value for age is given with respect to each yearly increase and for sex with respect to males. The heritability is reported as a fraction, multiplying this number by 100 will generate the heritability as a %.

The GWAS results were imported into Haploview [322] and are displayed as a Manhattan plot for each trait (Figure 2.6.a-d). SNPs with a p-value of less than 1E-05 are plotted above a blue line and are considered to show suggestive association with the Aβ trait. SNPs that achieve genome-wide significance with a p-value < 1.62E-07 are plotted above the red line. None of the SNPs achieved genome-wide significance in the study of Vis for any of the four traits evaluated. Eight SNPs achieved suggestive association with a p-value of <1E-05. Table 2.7 summarizes the "top-hits" for each of the four traits in the Vis GWAS analysis.

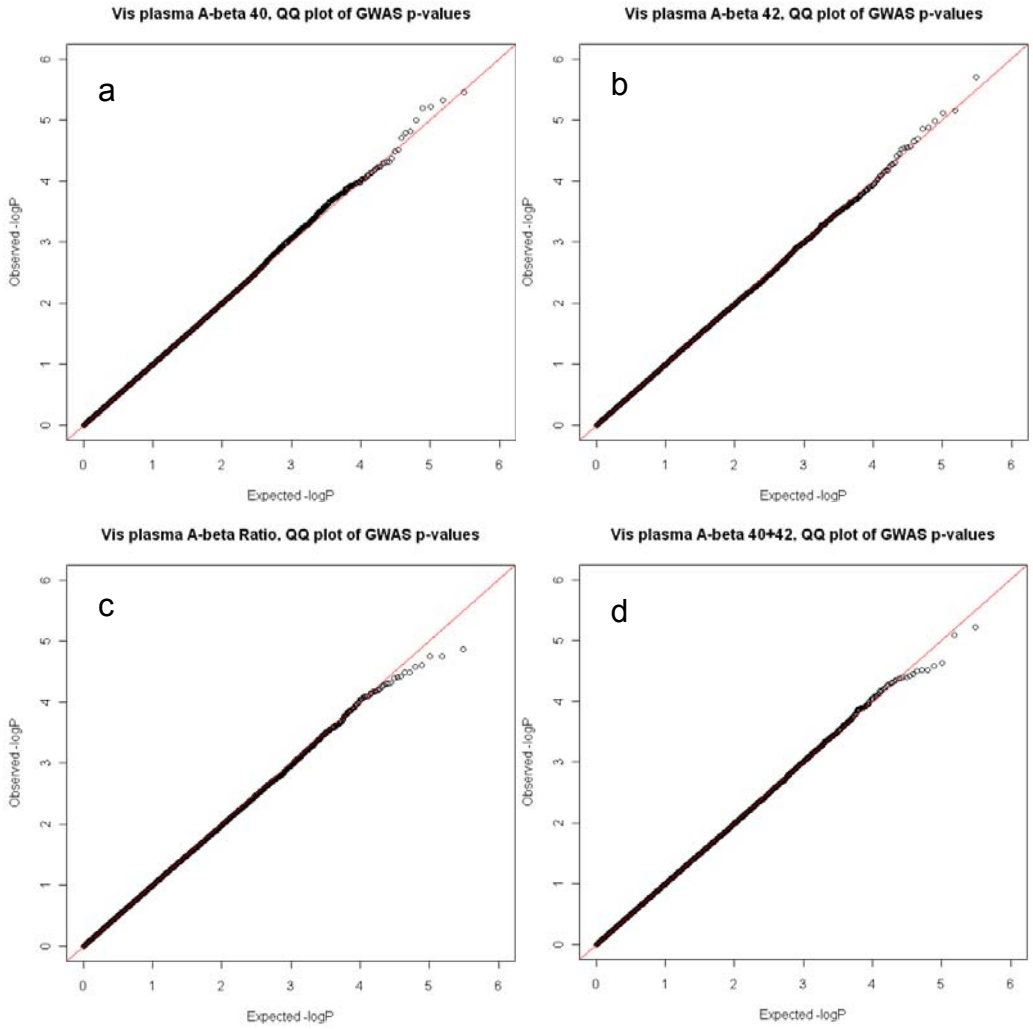

Figure 2.5. Quantile Quantile plots (QQ plots) of the Vis GWAS results on Aβ. QQ plots representing the distribution of observed p-values against the distribution of p-values as expected under the null hypothesis for the Vis GWAS of 308,998 SNPs for each of the plasma Aβ traits studied.

Trait	Chr	SNP	Position (bp)[a]	Allele	MAF	N	Beta	SE beta	p-value	Gene[b]	Location[c]
Aβ40	5	rs1422104	85,483,598	A	0.39	882	0.21	0.044	3.49E-06	NBPF22P	130kb, 5'
Aβ40	13	rs831165	110,590,297	A	0.31	882	0.22	0.049	4.78E-06	ARHGEF7	Intron 1
Aβ40	7	rs10250326	14,970,577	A	0.21	879	-0.25	0.056	6.04E-06	TMEM195	43kb, 5'
Aβ40	13	rs860520	110,584,006	A	0.31	883	0.22	0.049	6.39E-06	ARHGEF7	Intron 1
Aβ42	18	rs9952705	30,017,795	A	0.46	883	-0.23	0.049	1.99E-06	NOL4	Intron 1
Aβ42	18	rs4584903	29,972,523	G	0.42	882	-0.22	0.049	7.01E-06	NOL4	Intron 1
Aβ42	18	rs4239385	29,914,113	C	0.43	881	0.22	0.049	7.55E-06	NOL4	Intron 4
Aβ42+40	5	rs1422104	85,483,598	C	0.39	882	0.21	0.046	6.02E-06	NBPF22P	130kb, 5'
Aβ42+40	5	rs10035841	85,455,353	G	0.24	873	0.24	0.054	8.22E-06	NBPF22P	7kb, 5'

Table 2.7. Vis GWAS results with a p-value < 1E-05. Results are shown for three of the traits studied, the Aβ42:40 ratio did not identify any SNPs with a p-value less than 1E-05. **a**. Base pair positions based upon NCBI build 36 (March 2006), **b**. Gene listed is the closest gene in base pairs within 200kbp of the SNP. **c.** Location of SNP within a gene, or distance from and orientation with respect to the gene along the positive strand of the chromosome.

The results shown in Table 2.7 include the closest annotated gene to each of the associated SNPs. It is possible that the identified SNPs do not influence the gene to which they are physically closest so caution should be taken not to over interpret these results. Each of the genes listed in Table 2.7 are however briefly described below.

NBPF22P: neuroblastoma breakpoint family, member 22 (pseudogene). A literature search for this gene failed to yield any results, however the name of the gene implies a relationship to other cancer-related genes. It is difficult, at this time, to hypothesise as to how this locus might relate to plasma Aβ levels. Interestingly this gene does fall within the support interval for the Aß40 linkage peak identified on chromosome 5 in Table 2.5, and it is the only SNP with a suggestive p-value that falls within a region of suggestive linkage.

***ARHGEF7*:** Also known as COOL-1 or KIAA0142, ARHGEF7 is a Rho guanine nucleotide exchange factor. The literature indicates that the enzyme is required for interaction between p-21 activated kinases or PAKS and Rho family GTPases such as Cdc42 [348, 349]. The possible impact of SNPs at this locus with respect to plasma Aβ40 levels are not immediately clear, although the impact may affect downstream signalling.

TMEM195: Variants at the *TMEM195* locus have been reported to associate with type 2 diabetes mellitus (T2D) and/or glycemic traits in two large studies [316, 350]. SNP rs2191349 in the intergenic region between *TMEM195* and *DGKB* is reported to show association with both diabetes-related quantitative traits and the dichotomous T2D trait in the recent Dupuis *et al*. study [316]. Several studies have implicated a link between T2D and AD, some of which were most recently summarised by Han *et al*. [351]. In this context, the above result is of interest, but further study would be needed to elucidate the proposed association with LOAD.

***NOL4*:** Nucleolar protein four is a 524 residue protein that is expressed in brain and contains nuclear and nucleolar localization signal domains. There is little published literature on this protein [352, 353] and so it is difficult to build a hypothesis as to how variation at this locus might influence circulating Aβ levels.

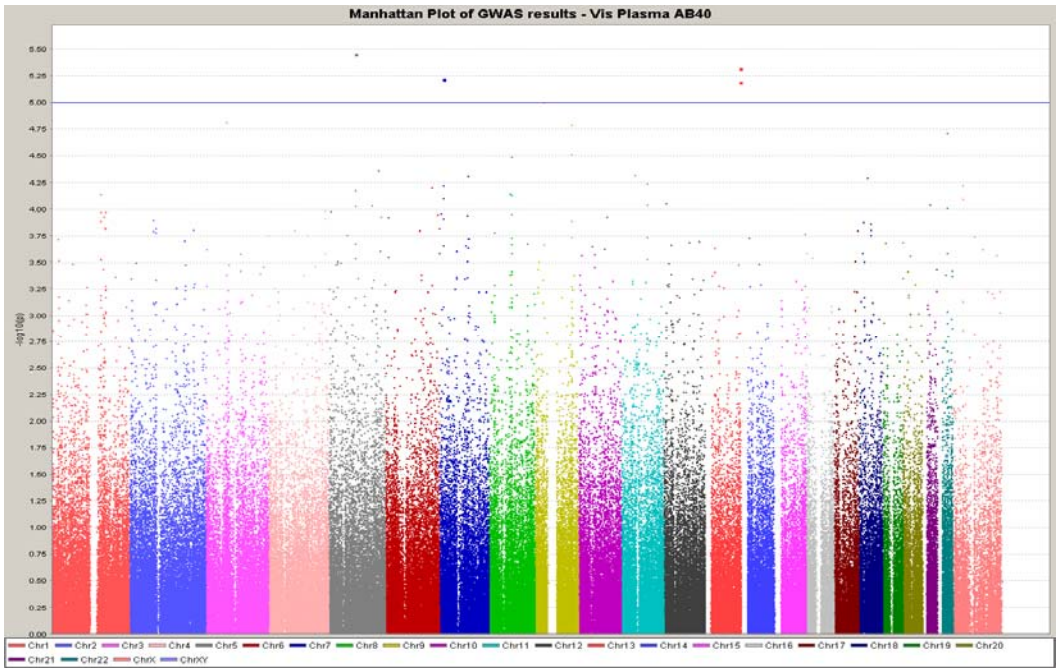

Figure 2.6a. Manhattan plot of GWAS results from plasma Aβ40 levels in Vis. Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

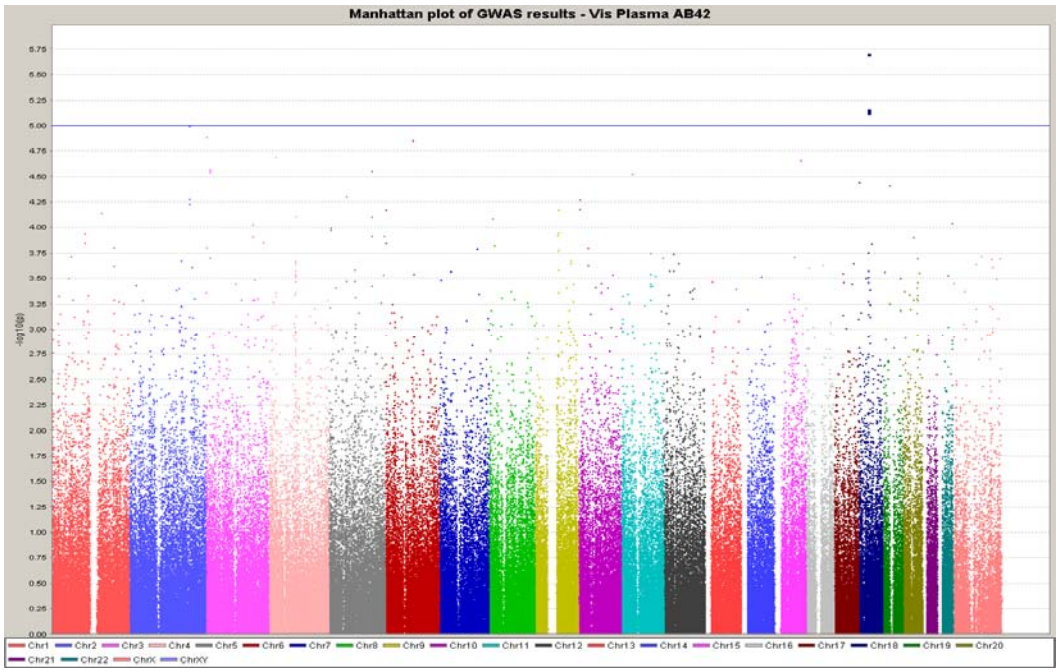

Figure 2.6b. Manhattan plot of GWAS results from plasma Aβ42 levels in Vis. Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

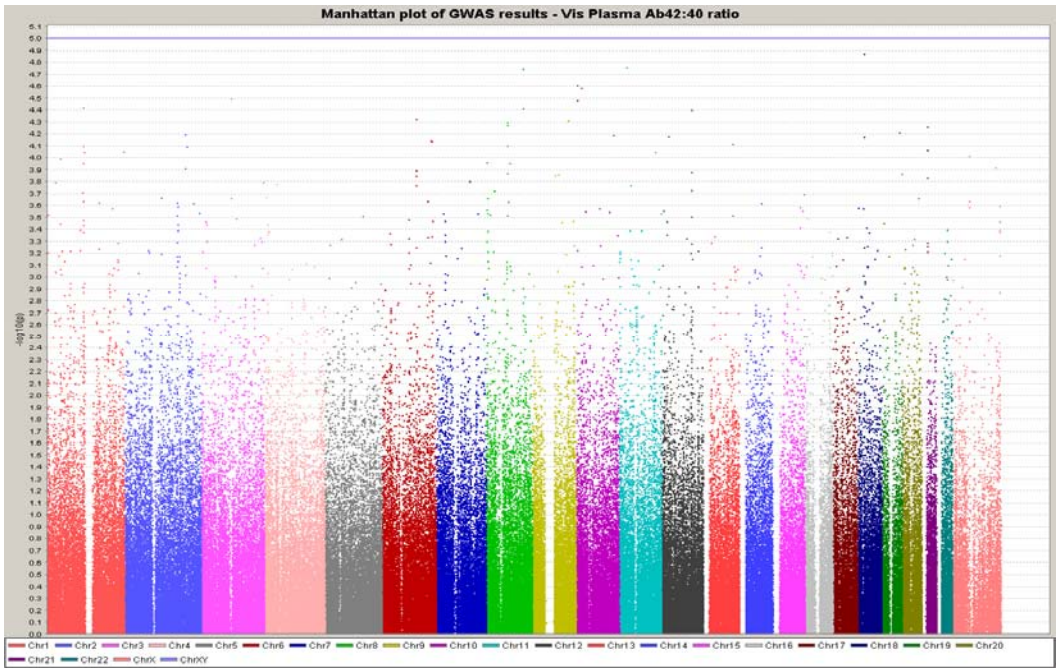

Figure 2.6c. Manhattan plot of GWAS results from plasma Aβ42:40 ratio measurements in Vis. Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

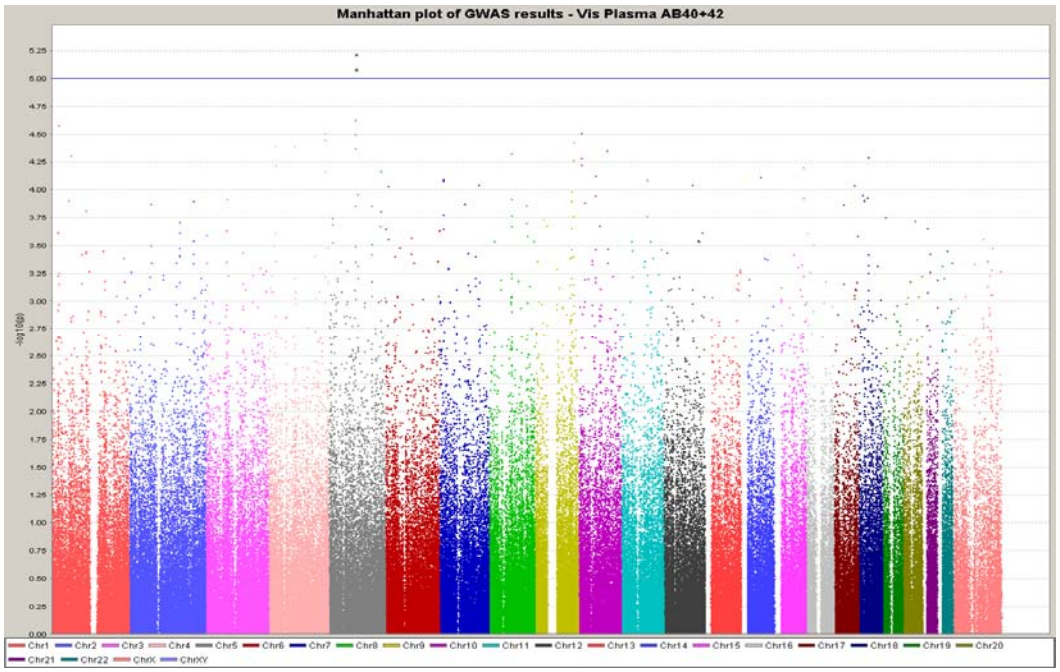

Figure 2.6d. Manhattan plot of GWAS results from plasma Aβ40+42 measurements in Vis. Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

2.3.4 Korcula GWAS

Quality Control: The quality control procedures for Korcula were run multiple times in succession, as previously performed for Vis. Prior to QC, there were 346,027 SNPs and 953 subjects. After QC procedures, there were 325,224 SNPs and 898 subjects. The majority of SNPs were removed due to low call rates (11,533 SNPs), with the other major reason being a minor allele frequency of less than 2% (10,384 SNPs). Some SNPs fell under both exclusion categories. The most common reason for removal of subjects from the analysis was likewise low call rate (39 subjects), with a small number being removed due to high autosomal heterozygosity, or IBS >95%, implying that the subject is either duplicated in the study or has a twin. In the QC procedure, a lenient p-value threshold of 1E-10 for deviation from Hardy Weinberg equilibrium was used for the same reasons as described for Vis. Thresholds of 95% and 97% for genotyping success rate for SNPs and subjects respectively was used. Following implementation of these QC thresholds, 98% of subjects and 97% of SNPs remaining had a proportion of successful genotypes greater than 99%; this is somewhat higher that the quality of calls found for Vis and may be a reflection on the quality of input DNA or differences in the quality of the genotyping chip used. Vis used the HumanHap300 (Ilumina) and Korcula used the HumanHap370CNV chip (Illumina). The mean heterozygosity per SNP is 0.35 and per subject is 0.34, again similar to Vis.

Polygenic model: The heritability of each trait, and the trait variance associated with covariates was obtained using the polygenic function available in GenABEL [318]. As described in Materials and Methods (section 2.2.8) the MLE for each of the

covariates and heritability, together with the corresponding p-values was also calculated. The *APOE ε4* covariate (presence/absence) was not nominally significant in any of the analyses ($p>0.05$) and so this covariate was not included in the polygenic model. This is a similar result to that observed in Vis for *APOE*. Consequently, only age and sex were included as covariates in the final polygenic models. The results for covariate and heritability estimation for each trait are shown in Table 2.8.

The plasma Aβ40 trait and the dependent traits of the ratio and total Aβ were highly significantly associated with increasing age. When this result is compared with the results for the same analysis in Vis, several similarities can be observed. The beta coefficient (effect size) for age on plasma Aβ40 in Vis was 0.025 and in Korcula it was 0.022. This result is very similar despite the different capture antibodies used in the two studies for measuring plasma Aβ40. Likewise the beta coefficients for the effect of age on plasma Aβ42 in the two populations was small and non-significant, although in Vis we observed a small positive beta coefficient (0.003) and in Korcula we observed a small negative beta coefficient (-0.004). Furthermore, the sex covariate did not show significant association with any of the plasma traits, similar to the result observed in Vis.

Korcula	Age		Sex		Heritability (h^2)	
Trait	beta	P-value	beta	P-value	beta	P-value
Plasma Aβ40	0.022	<0.0001	-0.024	0.849	0.467	<0.0001
Plasma Aβ42	-0.004	0.071	-0.100	0.159	0.197	0.030
Plasma Aβ42:40 ratio	-0.018	<0.0001	-0.075	0.257	0.175	0.030
Sum of plasma Aβ40+42	0.017	<0.0001	-0.047	0.582	0.421	<0.0001

Table 2.8. MLE of the effect size (beta coefficient) and statistical significance (p-value) in the Korcula population for covariates and heritability of Aβ quantitative traits. The estimates were obtained using the polygenic function in GenABEL. The beta coefficient for age is given per yearly increase and for sex the beta coefficient is shown with respect to males. The heritability is reported as a fraction, multiplying this number by 100 will generate the heritability as a %.

The mmscore test described and introduced by Chen WM and Abecasis G in 2007 [354], which implements FASTA (Family Score Test for Association) methods for analysis of the data, was used for genome wide association analysis of the 325,224 SNPs in 898 subjects from Korcula. The parameter λ which is an estimation of test statistic inflation (or deflation) used for genomic control methods can also be estimated using mmscore. In this analysis, λ is equal to between 0.96 and 1.00 and thus exhibits no requirement for implementation of genomic control. In fact, for 2 traits, λ indicates a deflation factor and implying that the test statistic may be underestimated. This can be observed graphically in the QQ plots (Figure 2.7 a-d), where the observed distribution falls below the red line which represents the distribution of p-values under the null hypothesis.

The GWAS results were imported into Haploview [322] and are displayed as a Manhattan plot for each trait. SNPs with a p-value of less than 1E-05 are plotted

above a blue line and are considered to show suggestive association. SNPs that achieve genome-wide significance with a p-value of < 1.54E-07 are plotted above the red line. None of the SNPs achieved genome-wide significance in the Korcula study for any of the four Aβ traits evaluated. Six SNPs achieved suggestive association, with a p-value of < 1E-05. Table 2.9 summarizes the "top-hits" for each of the four traits in the Korcula GWAS analysis.

Trait	Chr	SNP	Position (bp)[a]	Allele	MAF	N	Beta	SE beta	p- value	Gene[b]	Location[c]
Aβ40	21	rs2832643	30,450,088	G	0.23	875	-0.27	0.056	1.57E-06	CLDN17	10kb, 5'
Aβ40	21	rs2835933	38,045,624	G	0.19	876	0.29	0.060	1.78E-06	KCNJ6	Intron 2
Aβ42	6	rs6939603	5,792,055	G	0.26	872	-0.28	0.057	5.96E-07	FARS2	75kb, 3'
Aβ42	8	rs10505470	127,433,877	G	0.23	875	0.26	0.057	7.28E-06	FAM84B	200kb, 5'
Aβ42:40 Ratio	15	rs1668547	40,130,319	A	0.22	876	-0.28	0.057	1.31E-06	PLA2G4D	17kb, 5'
Aβ42:40 Ratio	1	rs12145722	22,615,572	G	0.27	875	0.27	0.062	9.06E-06	ZBTB40	35kb, 5'
Aβ42+40	21	rs2832643	30,450,088	G	0.23	875	-0.28	0.057	9.47E-07	CLDN17	10kb, 5'
Aβ42+40	6	rs6939603	5,792,055	G	0.26	872	-0.27	0.056	1.42E-06	FARS2	75kb, 3'
Aβ42+40	21	rs2835933	38,045,624	G	0.19	876	0.29	0.061	2.37E-06	KCNJ6	Intron 2

Table 2.9. Korcula GWAS results with a p-value < 1E-05. Results are shown for all four of the Aβ traits studied, **a**. Base pair positions based upon NCBI build 36 (March 2006), **b**. The genes listed are the closest genes in base pairs, within 200kbp of the SNP. **c.** Location of SNP within a gene or distance from gene and orientation with respect to the gene along the positive strand of the chromosome.

As mentioned previously in the GWAS analysis of Vis, caution must be taken not to over interpret the results with respect to the proximity of the identified SNPs and the genes listed in the Table. The genes listed are defined as those closest to the associated SNP within a 200kbp window, but they are frequently not the only gene within this window.

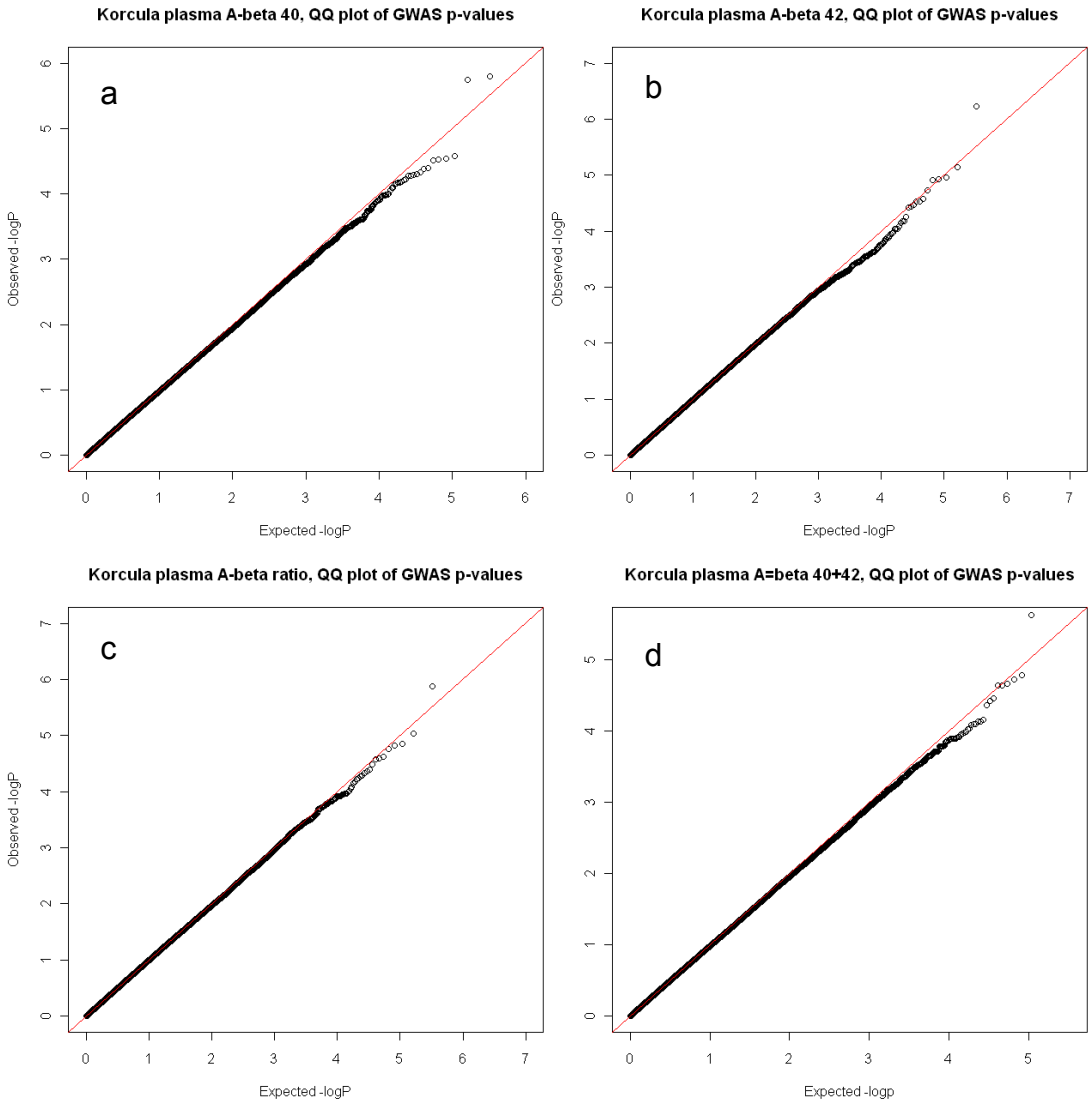

Figure 2.7 a-d. Quantile-Quantile plots (QQ plots). QQ plots representing the distribution of observed p-values against the distribution of p-values expected under the null hypothesis for the Korcula GWAS of 325,224 SNPs for each of the plasma Aβ traits studied.

CLDN17: The SNP at this locus is present 10kb 5' of the *CLDN17* gene with reference to build 36 of the human genome. However, the *CLDN17* gene is transcribed 3' to 5' and so this SNP is 3' of the *CLDN17* gene. *CLDN17* is one of a family of 23 Claudin genes [355]. There is little literature on the *CLDN17* gene, however the family of Claudin genes are known to be involved in the formation of tight cell junctions and have been implicated in cancers [356, 357]. With respect to Alzheimer's disease, the presence of this gene on chromosome 21 is of interest. The reasons for the relevance of chromosome 21 in Aβ and LOAD aetiology are discussed in chapter 1.

***KCNJ6*:** Also known as *GIRK2*, this gene encodes an inward rectifying ATP dependant potassium gated channel expressed in pancreatic beta cells and the brain. This gene also maps to chromosome 21. A mouse model, most commonly known as the *weaver* mouse, was generated by mutation of the mouse *GIRK2* gene and exhibits an ataxic phenotype [358]. Studies have indicated a role for *GIRK2* in long term potentiation and synaptic plasticity [359, 360], whilst two association studies investigating SNP's at this locus report one positive and one negative association with Alzheimer's disease [361, 362].

FARS2: This gene encodes a mitochondrial phenylalanyl-tRNA synthetase (PheRS) [363]. It is possible that variation at this locus could play a role in mitochondrial dysfunction which is a hypothesised pathogenic pathway for LOAD. However, there is limited literature on this gene and so a biological role in disease is difficult to hypothesise without additional information.

FAM84B: This gene maps to chromosome 8, encodes a protein that is a binding partner of alpha-1 catenin (*CTNNA1*) and is found at high levels in cancer cells [364]. The SNP on chromosome 8 that is proximal to *FAM84B* lies within a "gene desert" on chromosome 8 that has been linked to multiple cancers and is flanked by *FAM84B* and *C-MYC* [365]. It is unclear how this locus might relate to plasma Aβ or Alzheimer's disease other than the observation that the protein was found to bind *CTNNA1*. A related gene, *CTNNA3* lies within the 1 LOD support interval of the plasma Aβ42 linkage peak on chromosome 10 that was described by Ertekin-Taner *et al*. [62] and contains SNPs that have been reported to associate with both plasma Aβ and LOAD [206, 366].

PLA2G4D: This gene maps to chromosome 15 and is up-regulated in the skin from psoriasis patients and other inflammatory skin diseases [367]. Is not known to be expressed in the brain although SNPs at this locus have been investigated for association with schizophrenia [368]. A biological role for this locus in controlling plasma Aβ traits or LOAD susceptibility is not immediately clear.

ZBTB40: This gene is expressed in bone and two recent studies have implicated SNPs at this locus as being associated with bone mineral density, indicating a possible role in osteoporosis disease susceptibility [369, 370]. Again, there is no obvious biological role for the gene in controlling plasma Aβ traits.

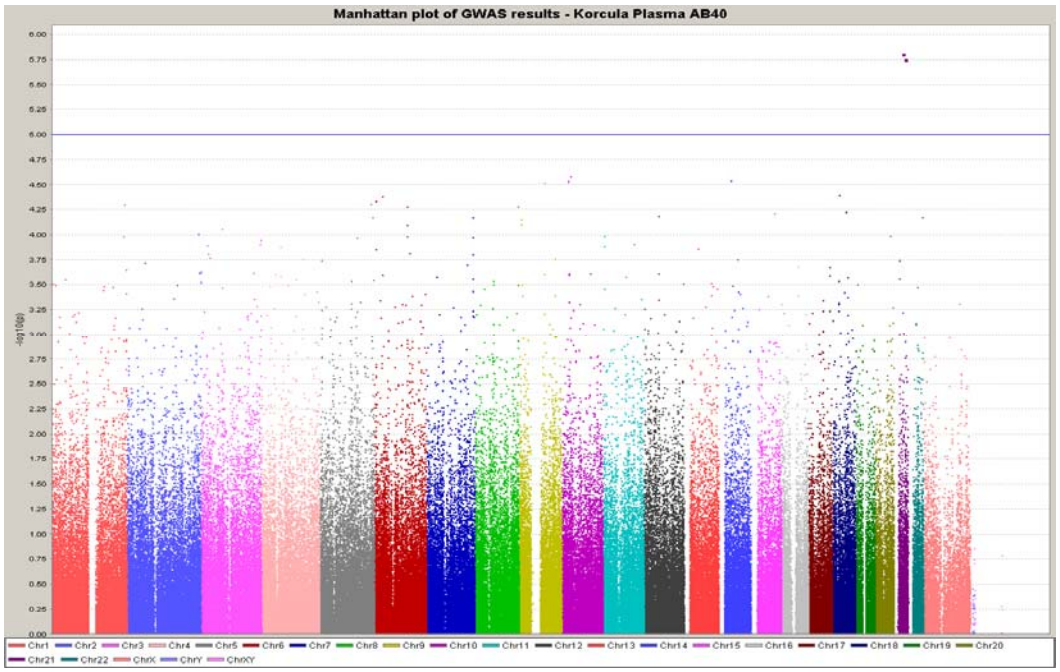

Figure 2.8a. Manhattan plot of GWAS results for plasma Aβ40 concentration in Korcula. Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

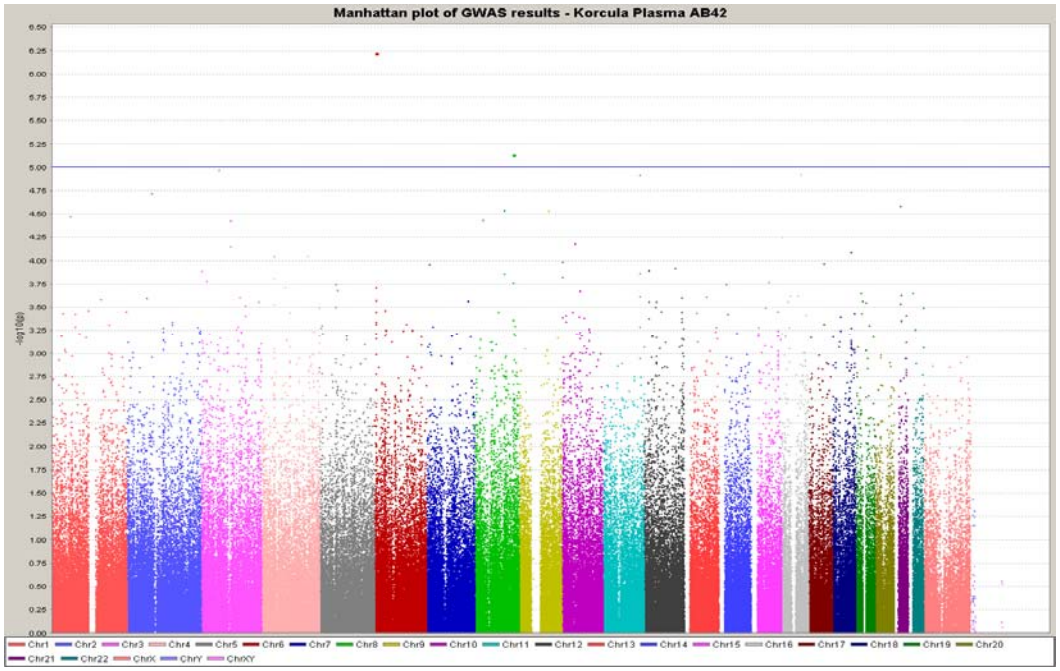

Figure 2.8b. Manhattan plot of GWAS results for plasma Aβ42 concentration in Korcula: Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

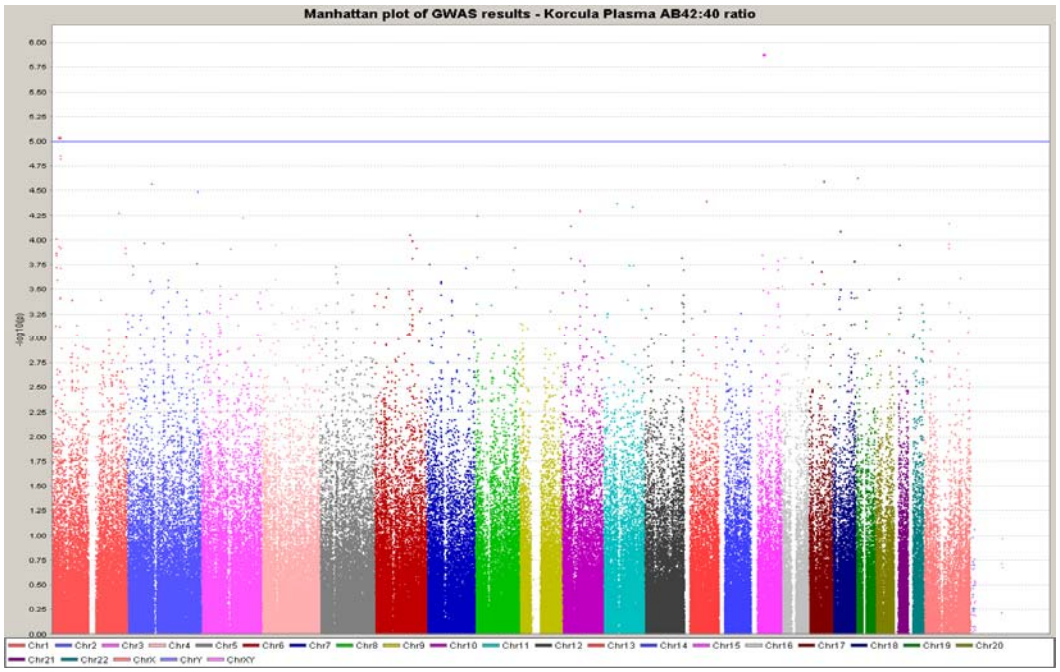

Figure 2.8c. Manhattan plot of GWAS results for plasma Aβ42:40 ratio measurements in Korcula: Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

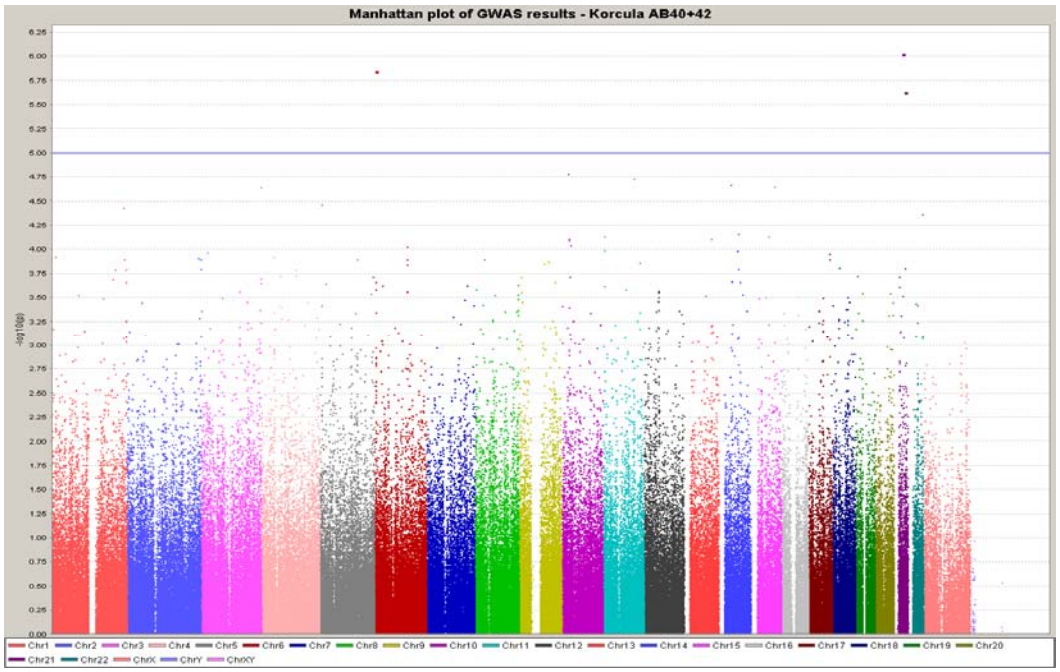

Figure 2.8d. Manhattan plot of GWAS results for plasma Aβ40+42 measurements in Korcula: Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The Blue horizontal line represents a p-value of 1E-05.

2.3.5 Meta-analysis

The R package MetABEL [318] was used to facilitate meta-analysis of the GWAS results generated from the Vis and Korcula studies. Meta-analysis was conducted for each of the four plasma Aβ traits and the results were reported for all SNPs that were present in both studies; in total 297,303 SNPs were present for meta-analysis in both studies. The results were imported into Haploview and are displayed as Manhattan plots. Table 2.10 outline the top hits.

Trait	Chr	SNP	Position (bp)[a]	Allele	MAF	N	Beta	SE beta	p- value	Gene[b]	Location[c]
Aβ40	6	rs11153495	115,108,714	G	0.18	1754	0.19	0.042	5.09E-06	-	-
Aβ42	X	rs6647514	73,802,014	A	0.23	1732	0.16	0.035	2.47E-06	KIAA2022	68kb, 5'
Aβ42	3	rs883370	133,109,978	G	0.39	1759	0.16	0.035	2.51E-06	CPNE4	Intron 1
Aβ42	14	rs12433867	92,134,836	G	0.18	1754	-0.25	0.055	5.45E-06	RIN3	Intron 3
Aβ42:40 Ratio	1	rs17521587	217,883,729	A	0.14	1758	-0.22	0.047	5.11E-06	-	-
Aβ42:40 Ratio	2	rs4848444	117,954,791	A	0.18	1758	0.19	0.043	6.57E-06	-	-
Aβ42+40	12	rs12320234	76,427,966	A	0.04	1759	-0.36	0.081	8.63E-06	-	-

Table 2.10. Meta-analysis of GWAS results for Aβ traits with a p-value <1E-05. Results are shown for all four of the traits studied. **a**. Base pair positions based upon NCBI build 36 (March 2006), **b**. Gene listed is the closest gene in base pairs within 200kbp of the SNP. **c.** Location of SNP within a gene or distance from gene and orientation with respect to the gene along the positive strand of the chromosome.

KIAA2022: A literature search for this gene reveals that it is both highly expressed in brain and has been shown to be associated with mental retardation in males when the gene is disrupted [371]. Too little is known however regarding the gene to hypothesise a possible biological role in plasma Aβ trait control or LOAD association.

CPNE4: The gene is also known as copine IV and enocodes a calcium dependant membrane binding protein. Very little has been published on this gene or the protein it encodes although more is known about the related protein *CPNE1*. It is likely that the family of copines play a role in signal transduction and there is evidence for the expression of some members of this family in the brain [372, 373].

RIN3: is a widely expressed gene encoding Ras interaction-interference protein (Rin). It belongs to a large family of Rin proteins which are known to be binding partners to the RAB5 small GTPases. Evidence reported in the literature indicates that this protein plays a role in intracellular transport, particularly indicating a role in stabilising Rab5 and thus the endocytic transport pathway from the plasma membrane to early endosomes [374]. It may be hypothesised that RIN3 could play a role in APP processing via the endocytic pathway; however more research would be necessary to investigate this hypothesis.

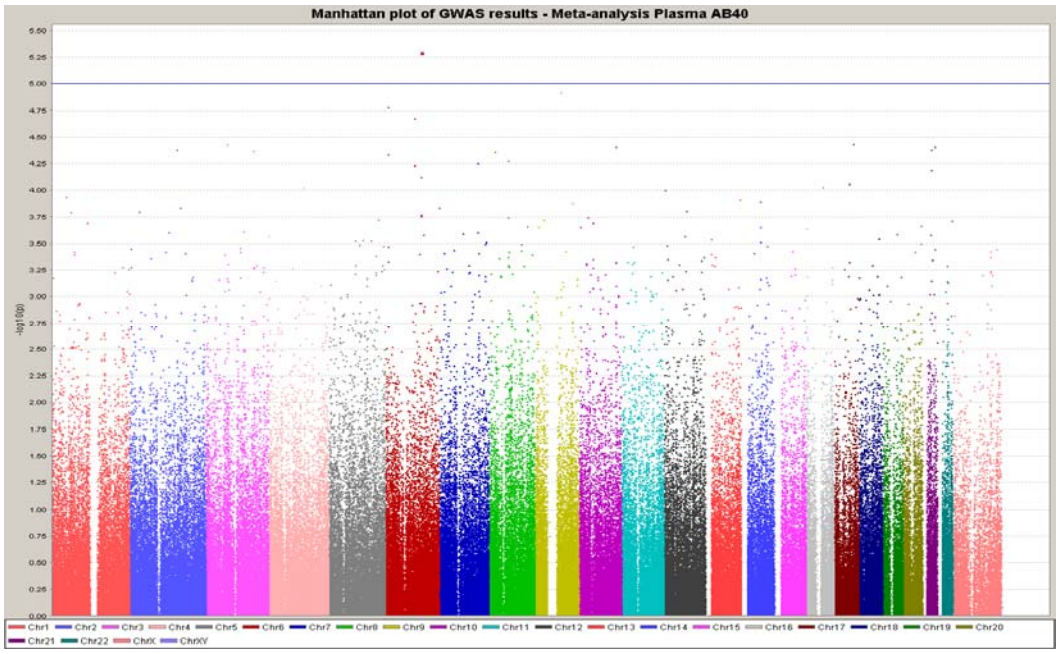

Figure 2.9a. Manhattan plot of GWAS results from meta-analysis of Vis and Korcula plasma Aβ40 concentration: Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

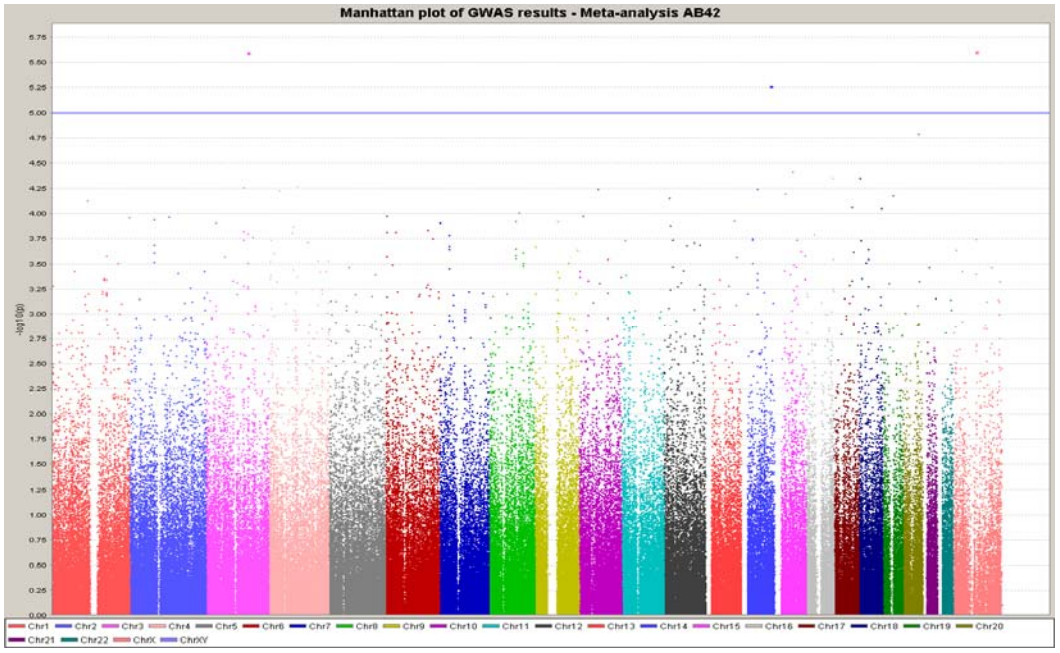

Figure 2.9b. Manhattan plot of GWAS results from meta-analysis of Vis and Korcula plasma Aβ42 concentrations: Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

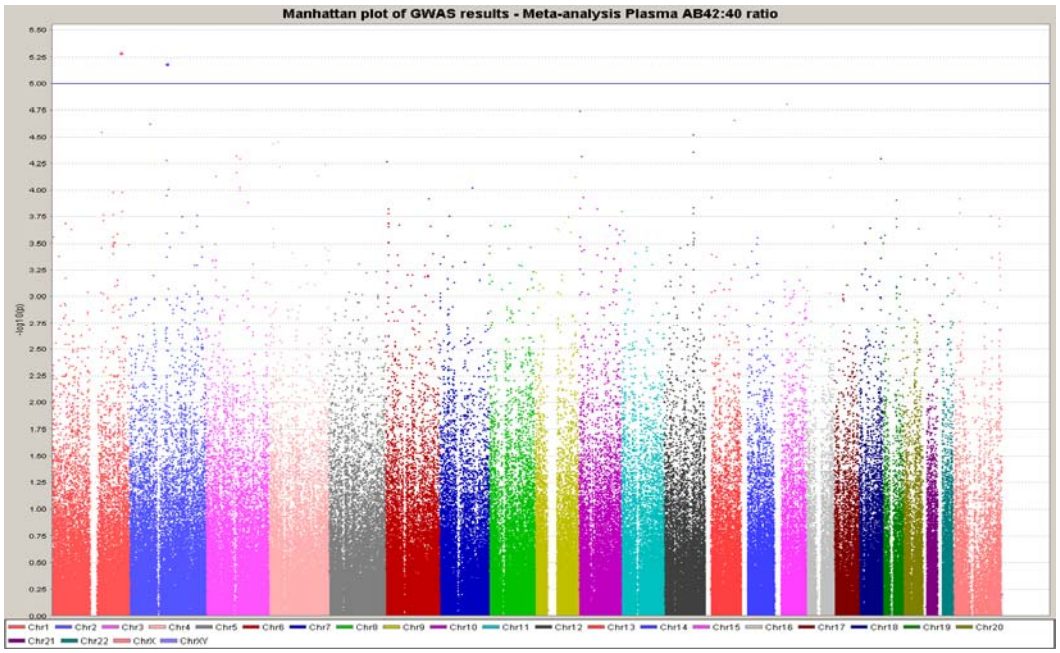

Figure 2.9c. Manhattan plot of GWAS results from meta-analysis of Vis and Korcula plasma Aβ42:40 ratio measurements. Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

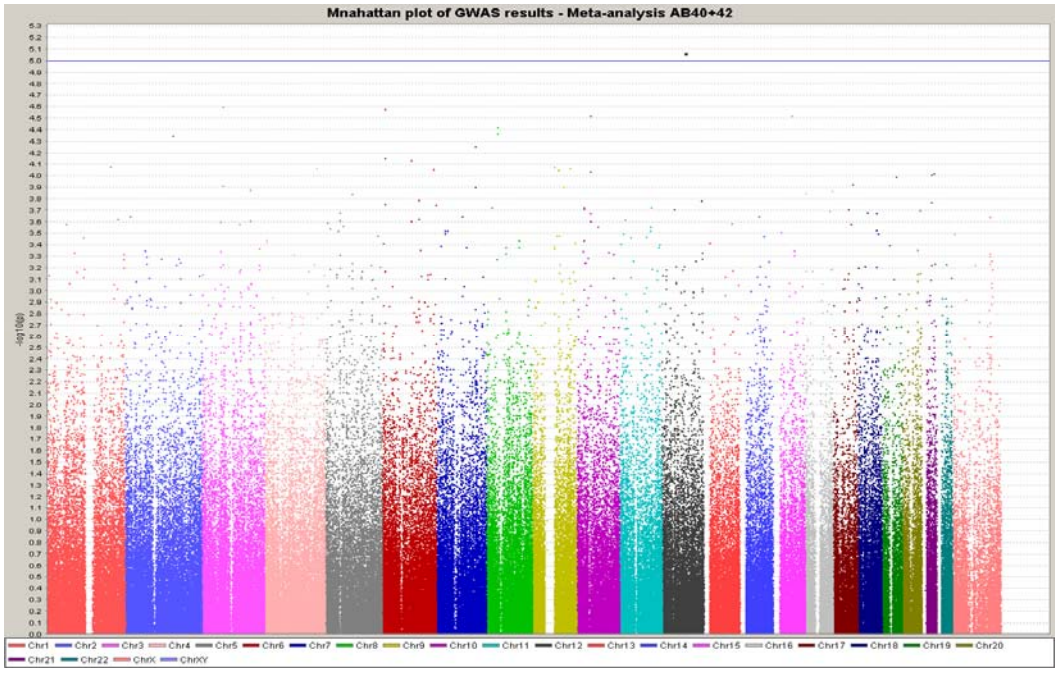

Figure 2.9d. Manhattan plot of GWAS results from meta-analysis of Vis and Korcula plasma Aβ40+42 measurements: Chromosomes are plotted on the X axis and $-\log_{10}$ adjusted p-values are plotted along the Y axis. The blue horizontal line represents a p-value of 1E-05.

2.3.6 Association of top plasma Aβ SNPs with LOAD

The aim of this chapter was to identify candidate LOAD loci through the use of GWAS, endophenotypes and isolated population based approaches. Using these methods, 20 SNPs were identified to show suggestive association with plasma Aβ (tables 2.7, 2.9 and 2.10). Of these, 19 were genotyped as part of a LOAD GWAS published in 2009 (Carrasquillo et al). A query of these LOAD GWAS results, for the 19 SNPs identified, is presented in Table 2.11.

One SNP, rs12320234, achieved nominally significant association with LOAD (OR = 1.46, p = 0.027), however this would not maintain significance after bonferonni correction for multiple tests.

An evaluation of 19 SNPs would be expected, by chance, to identify 2 SNPs with a p-value of less than 0.1. In this study, in addition to rs12320234, a further five SNPs achieved a p-value of less than 0.1 (rs12145722, rs17521587, rs12433867, rs4239385 and rs2835933), resulting in a far greater number of SNPs with a p-value of less than 0.1 than would be expected by chance. Whilst these SNPs do not individually achieve significant association with LOAD, it can be argued that these SNPs are feasible candidates for follow up with additional genotyping in a greater number of LOAD cases and controls.

CHR	SNP	Position (bp)	Allele	MAF (N) Cases	MAF (N) Controls	OR	95% CI	p-value
1	rs12145722	22,615,572	G	0.19 (733)	0.17 (1103)	1.21	1.00-1.47	0.055
1	rs17521587	217,883,729	A	0.13 (733)	0.15 (1104)	0.83	0.67-1.04	0.099
2	rs4848444	117,954,791	A	0.15 (732)	0.14 (1103)	1.13	0.92-1.40	0.253
3	rs883370	133,109,978	G	0.44 (733)	0.44 (1102)	0.94	0.81-1.09	0.424
5	rs10035841	85,455,353	A	0.34 (729)	0.32 (1096)	1.02	0.87-1.19	0.813
5	rs1422104	85,483,598	A	0.48 (733)	0.47 (1103)	0.98	0.85-1.14	0.834
6	rs11153495	115,108,714	G	0.18 (732)	0.18 (1100)	1.04	0.86-1.27	0.658
8	rs10505470	127,433,877	A	0.22 (733)	0.22 (1103)	1.00	0.84-1.20	0.972
12	rs12320234	76,427,966	A	0.06 (733)	0.04 (1105)	1.46	1.04-2.04	0.027
13	rs860520	110,584,006	A	0.29 (732)	0.30 (1105)	0.92	0.78-1.08	0.289
13	rs831165	110,590,297	A	0.29 (733)	0.30 (1105)	0.91	0.78-1.07	0.268
14	rs12433867	92,134,836	G	0.08 (730)	0.11 (1099)	0.78	0.60-1.01	0.056
15	rs1668547	40,130,319	A	0.27 (733)	0.27 (1102)	1.04	0.88-1.23	0.655
18	rs4239385	29,914,113	C	0.41 (730)	0.45 (1102)	0.87	0.75-1.01	0.074
18	rs4584903	29,972,523	G	0.44 (732)	0.44 (1101)	1.05	0.90-1.22	0.541
18	rs9952705	30,017,795	A	0.48 (733)	0.47 (1104)	1.06	0.91-1.23	0.442
21	rs2832643	30,450,088	G	0.26 (730)	0.23 (1101)	1.08	0.91-1.29	0.385
21	rs2835933	38,045,624	G	0.19 (732)	0.22 (1104)	0.84	0.70-1.01	0.061
23	rs6647514	73,802,014	A	0.19 (578)	0.19 (835)	1.04	0.84-1.30	0.722

Table 2.11 Association analysis of 19 SNPs with LOAD. SNPs identified as the top candidates in the plasma Aβ GWAS described in this chapter are tested for association with LOAD. MAF = Minor Allele Frequency, CHR = chromosome.

2.3.7 Association of LOAD SNPs with plasma Aβ

Recent LOAD GWAS have identified replicable genome-wide significant association for SNPs at three loci frequently identified by the proximal candidate genes *CLU*, *CR1* and *PICALM*. Of the reported SNPs at these loci, three were present as part of the plasma Aβ GWAS described in this chapter and were investigated for association with plasma Aβ levels in both Vis and Korcula using the methods described previously. The results are presented in Table 2.12.

This analysis identified nominally significant association with increased plasma Aβ42 levels for the SNP at the PICALM locus (rs3851179), in the Korcula population (p=0.04) and in the meta analysis of both Vis and Korcula (p=0.027), however these p-values would not maintain significance after correction for multiple tests.

Although these results indicate a suggestive association for the PICALM locus with plasma Aβ levels, the p-values reported for association with LOAD are much smaller and so it is unlikely that the association observed at these loci influences risk for LOAD via molecular pathways that likewise influence plasma levels of Aβ.

CHR	SNP	Position (bp)	Gene	Allele	MAF			Trait	Vis		Korcula		Meta	
					Vis	Korcula	Meta		beta	p-value	beta	p-value	beta	p-value
8	rs11136000	27,520,436	CLU	A	0.36	0.37	0.36	A□40	0.01	0.753	-0.09	0.071	-0.03	0.311
								A□42	0.02	0.762	-0.07	0.139	-0.03	0.400
								Ab42/40 ratio	0.00	0.995	-0.02	0.631	-0.01	0.735
								Ab40+42	0.02	0.653	-0.10	0.053	-0.04	0.306
11	rs3851179	85,546,288	PICALM	G	0.31	0.36	0.34	A□40	0.04	0.453	0.05	0.260	0.04	0.185
								A□42	0.05	0.294	0.10	**0.040**	0.08	**0.027**
								Ab42/40 ratio	0.04	0.465	0.07	0.155	0.05	0.125
								Ab40+42	0.05	0.289	0.07	0.166	0.06	0.084
1	rs6701713	205,852,912	CR1	A	0.23	0.25	0.24	A□40	0.06	0.226	-0.01	0.818	0.03	0.473
								A□42	0.03	0.605	-0.04	0.477	-0.01	0.886
								Ab42/40 ratio	-0.04	0.502	-0.03	0.532	-0.04	0.359
								Ab40+42	0.06	0.238	-0.02	0.778	0.03	0.517

Table 2.12 Analysis of recently identified LOAD SNPs, for association with plasma Aβ levels. MAF = Minor Allele Frequency.

2.4 Summary

Assessment of the heritability of plasma Aβ concentrations in Vis and Korcula revealed surprising differences in these two populations, where Vis showed only weak evidence for heritability, conversely, plasma Aβ in Korcula was significantly heritable. A genome-wide linkage study of Vis identified several suggestive QTL, some of which overlap with known LOAD linkage peaks and provide both interesting and relevant candidates for further study as well as supportive evidence for functional variation at these loci. GWAS analysis, did not identify any QTL that achieved genome-wide significance according to Bonferroni correction standards for any of the traits in either study or by meta-analysis. However, several promising loci with low p-values that are good candidates for further study were identified. Nineteen of these loci were tested for association with LOAD and whilst significant association with LOAD, following correction for multiple tests, was not found for any of these SNPs there were many more SNPs with a p-value of less than 0.1 than would be expected by chance. Conversely, SNPs recently identified for association with LOAD did not exhibit significant association with plasma Aβ in these populations.

These studies have also provided data for future use in phenotype convergence studies where low p-values recorded for multiple traits such as plasma Aβ, LOAD phenotype and gene expression for a given SNP or gene may be informative in a situation where individual studies may be under-powered and heterogeneity complicates traditional replication studies. One example of this comes from two

simultaneously published studies of the insulin degrading enzyme [3, 375]. This will be discussed further in Chapter 3.

Chapter 3.

3.1 Background

In the previous chapter the use of isolated populations, quantitative traits and whole genome analyses were used as tools to identify loci that may play a role in LOAD by influencing levels of the Aβ peptide. In this chapter, the results of these studies are combined with gene expression traits and the LOAD disease trait in order to execute phenotypic convergence analysis. The purpose of this approach is to facilitate the identification of loci that may not achieve genome-wide significant association (α typically less than 1E-07) with any one trait, but nevertheless exhibit nominal evidence for multiple relevant traits. Whilst the loci identified using this approach likely have weak effects on the disease trait, it is becoming more widely accepted that complex traits with a substantial genetic component, such as LOAD, are likely caused by variation at multiple loci with individually weak effects. The identification of these additional loci, whilst challenging, is pertinent to improving our understanding of Alzheimer's disease. Since the majority of genes implicated in AD appear to have some involvement in Aß metabolism it is logical to evaluate the effect of SNPs in these genes on plasma Aß levels. In this chapter a phenotype convergence approach is used to evaluate the influence of SNPs at the IDE locus on expression of the IDE gene, levels of plasma Aβ and risk for LOAD.

3.1.1 Aβ degrading enzymes

As discussed in chapter one the amyloid cascade hypothesis postulates that the major triggering event in the development of Alzheimer's disease is likely increased levels of the Aβ peptide in the brain [137]. Changes in Aβ levels most likely occur through alterations of two main pathways; Aβ generation and Aβ degradation. Several enzymes are known to degrade Aß, among the most extensively studied are, the Insulin Degrading Enzyme (IDE), which degrades both insulin and Aß, Neprylisin (NEP) and the angiotensin converting enzyme (ACE) [205, 376-378]. Whilst the Aβ degrading properties of these enzymes are well established, their role in LOAD has been difficult to elucidate and genetic studies have produced conflicting and inconsistent results [31].

3.1.2 *IDE*.

The *IDE* locus is of particular interest in the study of LOAD since it is not only a functional candidate but is also proximal to a linkage peak identified through analysis of AD families [379]. For these reasons IDE was explored by the Younkin laboratory as a candidate LOAD gene through targeted genotyping of conserved variants (>70% sequence identity in human vs rodent) based upon the hypothesis that conserved variants have an increased likelihood to impact phenotypes. The initial target phenotype (or trait) in this study was expression levels of the *IDE* gene in the cerebellum of AD subjects. In post-mortem AD brains, AD pathology is usually observed in many areas of the brain. The cerebellum is unusual in that it is virtually unaffected and shows no dense amyloid plaques or NFT's. Thus the

cerebellum provides an opportunity to study gene expression levels in the target tissue (brain) whilst controlling for the downstream effects of the disease process as much as possible. IDE transcript concentration was chosen as a target trait based upon the hypothesis that variants that influence disease might do so by influencing the expression of certain genes relevant to the disease. In this chapter two parallel studies are highlighted that identified SNP's in conserved regions of the *IDE* locus that significantly associated with *IDE* expression levels in the AD cerebellum studied and likewise exhibited significant association with LOAD [3, 205, 375].

The most important of these SNPs was investigated for association with plasma levels of Aβ in the Vis and Korcula populations using the data that were described in the previous chapter. The association detected for this SNP, for all three traits, correlated in a logical manner such that the allele that was associated with increased IDE expression was likewise associated with decreased levels of plasma Aβ and decreased risk for LOAD [3]. This result is an example of the utility of the phenotypic convergence approach, to identify LOAD loci that would otherwise go undetected using whole genome analysis approaches for a single trait.

The results on Aß reported in the previous Chapter greatly strengthened the paper reporting these convergent results that was published in 2010 [3]. The author of this thesis was an equally contributing author on that paper, and was involved in writing and editing the manuscript. For this reason and to present rigorously the entire convergent set of results, sections from the text of the published manuscript have been liberally incorporated into parts of this chapter.

3.2 Methods

The methods described in this chapter are largely taken from Carrasquillo *et al* [3].

3.2.1 Subjects

Approval was obtained from the ethics committee or institutional review board of each institution responsible for the ascertainment and collection of samples (Mayo Clinic College of Medicine, Jacksonville, FL and Mayo Clinic College of Medicine, Rochester, MN, USA; MRC Human Genetics Unit, Western General Hospital and the University of Edinburgh Medical School, Edinburgh, UK; University of Split Medical School and University Hospital "Sestre Milosrdnice", Croatia; Queen's Medical Centre, Nottingham, UK; Queen's University Belfast, Northern Ireland, UK; University of Bristol, Frenchay Hospital, UK; University of Manchester, UK; University of Oxford, UK; and University of Bonn, Germany. Written informed consent was obtained for all individuals that participated in this study.

The age-at-diagnosis/entry, diagnosis method, and percent female in the LOAD cases and controls for each USA and ART series are described in Table 3.1.

3.2.2 DNA Isolation

For the JS and RS samples, DNA was isolated from whole blood using an AutoGen instrument (AutoGen, Inc, Holliston, MA). The DNA from AUT samples was extracted from cerebellum using Wizard® Genomic DNA Purification Kits (Promega Corp., Madison, WI). DNA from the RS and AUT series was scarce, so samples from these two series were subjected to whole genome amplification using

the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences Corp., Piscataway, NJ).

For the ART samples, genomic DNA was extracted from whole blood samples or brain tissue using the QIAamp DNA blood mini kit (Qiagen, Crawley, West Sussex, UK).

Series	AD	CTRL	Total	Mean Age (SD)		% Female	
				AD	CTRL	AD	CTRL
USA							
JS	632	660	1,292	78.1 (6.1)	77.5 (7.7)	62.3	61.4
RS	576	1,417	1,993	79.7 (7.9)	78.4 (6.0)	63.4	54.6
AUT	603	374	977	81.1 (8.6)	75.7 (8.3)	59.2	42.8
NCRAD	702	209	911	75.2 (6.8)	78.3 (8.9)	64.8	61.7
Total	2,513	2,660	5,173	78.4 (7.7)	77.8 (7.1)	62.5	55.2
ART							
Belfast	237	234	471	78.4 (6.7)	76.2 (7.5)	65.8	65.4
Bonn	175	199	374	75.9 (7.6)	71.0 (6.9)	76.0	49.2
Manchester	210	0	210	68.7 (6.3)	N/A	50.0	N/A
Oxford	167	205	372	73.0 (6.8)	77.3 (8.1)	55.7	57.1
Southampton	227	143	370	80.9 (7.5)	75.8 (6.6)	59.0	49.0
Total	1,016	781	1,797	75.6 (8.2)	75.1 (7.8)	61.1	56.1
USA and ART	3,529	3,441	6,970	77.6 (7.9)	77.2 (7.3)	62.1	55.4

Table 3.1. Subjects. The number of AD patients (AD) and controls (CTRL), mean age, and the percentage that are female are given for each individual and pooled series. Mean age is given as age at diagnosis/entry. The standard deviation (SD) from the mean is given in parenthesis

3.2.3 Variant Discovery

Conserved segments in the region encompassing all 25 exons of *IDE* and up to 10 kb from the 5′ and 3′ ends of the gene, were screened for DNA variation likely to be functional in a subset of the JS series (269 cases and 252 controls). This sample size provided enough power to detect all polymorphisms that occur at a frequency >1%

in this population. The criteria used for conservation was a cut-off of ≥70% identity over 100 bp windows between the human and mouse sequence or between the human and rat sequence, as determined by the pre-computed alignments in the VISTA Genome Browser (http://pipeline.lbl.gov/cgi-bin/gateway2). The human genome position screened was chr10:94,335,255-94,475,049 in the Human April 2003 genome build. The latter build was used as the base genome, and was compared to the Mouse Feb. 2003 and Rat Jan. 2003 builds. In each of the 521 subjects screened, 40 amplicons were evaluated that contained a total of 9,501 bp of conserved DNA. Forty PCR primer pairs were designed to screen the targeted conserved segments via denaturing high performance liquid chromatography (dHPLC). This variant discovery effort yielded 53 variants (42 novel) of which 14 had minor allele frequencies (MAFs) >1%. Twelve of these were genotyped, 2 could not be converted into TaqMan assays (deep intronic, 1 base insertions within simple repeats). An additional review of all known *IDE* variants identified 15 additional variants, each located in a 100 bp region that had sequence identity of ≥70% between the mouse and human sequence. Of these, 8 had MAFs≥1%. Three of these were in perfect LD with variants from the de novo search, thus were eliminated leaving 17 common *IDE* variants in conserved regions to be genotyped.

Additional details regarding variant discovery is detailed in Carrasquillo et al [3].

3.2.4 Genotyping of *IDE* Variants

All genotyping of USA series samples was performed at the Mayo Clinic in Jacksonville using TaqMan® SNP Genotyping Assays in an ABI PRISM® 7900HT

Sequence Detection System with 384-Well Block Module from Applied Biosystems, California, USA. The genotype data was analyzed using the SDS software version 2.2.2 (Applied Biosystems, California, USA).

The ART data, was also generated with TaqMan assays performed at Geneservice (Cambridge, UK).

3.2.5 *IDE* Haplotypes

Analysis by HaploView [322] (solid spine of LD) showed that the 17 *IDE* variants were in strong linkage disequilibrium.. These variants formed 11 haplotypes with a frequency of 1.0% or more that could be identified with 10 tagging variants. Analysis by Haplo.Stats identified the same 11 haplotypes, which accounted for 96% of the *IDE* genetic variation captured by the 17 variants. Haplotype frequencies were estimated using the expectation-maximization approach implemented in the haplo.em function of haplo.stats v1.2.2 [380]. Haplo.stats was also used to create files showing the inferred haplotypes for each individual, which were used to perform logistic regression analyses.

3.2.6 Measurement of *IDE* mRNA Expression

Total RNA was extracted from samples of cerebellum from 200 AD brains using an ABI PRISM 6100 Nucleic Acid PrepStation and the Total RNA Isolation Chemistry kit from Applied Biosystems. RNA was reverse transcribed to single-stranded cDNA using the High-Capacity cDNA Archive Kit from Applied Biosystems. Real-time quantitative PCR was performed in triplicate for each sample using ABI

TaqMan Low Density expression Arrays (384-Well Micro Fluidic Cards) with a pre-validated TaqMan Gene Expression Assay. 18s ribosomal RNA (18s rRNA) was used as the endogenous control for the relative quantification of *IDE* mRNA. Real-time PCR cycle threshold (C_T) raw data was collected and exported using the ABI PRISM® SDS software version 2.2. Additional details can be found at Carrasquillo et al [3]

3.2.7 Association of *IDE* Haplotypes and Variants with *IDE* Transcript

The fold difference in the expression levels for the minor allele for each variant was estimated from the regression coefficient (ß) for the additive model that was fit to the ΔC_T data, which were expressed on the $\log_2$ scale, while also adjusting for *APOE* ε4 dosage, age at diagnosis/entry and sex. Therefore, the associations were summarised using $2^{(-ß)}$ to estimate the fold difference in expression due to a single copy of the minor allele, and $2^{[-(ß \pm 2SEM)]}$ to estimate the corresponding 95% confidence interval (CI). The expression differences were further summarised graphically by preparing box plots showing the relative amount of *IDE* transcript in any sample in the series as compared to the mean expression level of all of the samples that were major homozygotes. This was done by computing the mean of all ΔC_T values for the samples in the major homozygote category and subtracting it from each of the ΔC_T values for all samples in all genotype categories (i.e. major homozygotes, heterozygotes and minor homozygotes) to obtain $\Delta\Delta C_T$ values. Given that these $\Delta\Delta C_T$ values are the $\log_2$ fold changes between two measurements, the relative expression levels between any sample and the mean of the major

homozygotes is given by $2^{-\Delta\Delta CT} = 2^{(-(\Delta C_{T\_Any\ Sample} - mean(\Delta C_{T\_Major\ Homozygotes})))}$. These $2^{-\Delta\Delta CT}$ values are plotted in the box plots.

The overall association between estimated haplotypes (frequency >0.01 in combined dataset) and the level of *IDE* transcript (ΔC_T) was determined by multivariable linear regression using a haplotypic dosage model with sex, age at diagnosis/entry, and *APOE* ε4 dosage (0, 1 or 2 copies of the ε4 alelle) as covariates. The association between individual haplotypes and *IDE* transcript (ΔC_T) was determined by linear regression using an allelic dosage model with sex, age at diagnosis/entry, and *APOE* ε4 dosage as covariates. The haplo.stats package was used to create the files containing the inferred haplotypes for each individual that were used to perform these analyses.

IDE variants were analyzed for association with the level of *IDE* transcript (ΔC_T) using an additive/allelic dosage (11 = 0, 12 = 1, 22 = 2) linear regression model that included sex, age at diagnosis/entry, *APOE* ε4 dosage and series as covariates.

3.2.8 Association of *IDE* Haplotypes and Variants with LOAD

The overall association between haplotypes (frequency >0.01 in combined dataset) and LOAD was determined using the global score statistic implemented in the haplo.score function incorporated in the haplo.stats package [380] while adjusting for sex, age at diagnosis/entry, *APOE* ε4 dosage, and series effects. The association between individual haplotypes and LOAD was determined by entering the estimated haplotypes for the cases and controls into univariate logistic regression using a

haplotypic dosage model that adjusted for sex, age at diagnosis/entry, *APOE* ε4 dosage, and series as covariates.

IDE variants were analyzed for association with LOAD by logistic regression using PLINK software [381] with sex, age at diagnosis/entry, and *APOE* ε4 dosage as covariates. For logistic regression using data from the combined series, we also included indicator variables for the individual series as covariates. All logistic regression analyses were performed using the additive/allelic dosage model as the primary test of significance for each variant. To test for evidence of any series-related effects on these associations, we compared this model to one that also incorporated series x genotype interactions in addition to all previously mentioned covariates. This analysis showed no evidence for series-related effects on any of the associations.

3.2.9 Linkage Disequilibrium between rs6583817 and rs7910977

A LOAD genome wide association study (GWAS) was previously conducted with 313,504 SNPs in 844 cases and 1,255 controls that were also analyzed for association with the 17 *IDE* variants in conserved regions [241]. HaploView 3.1 was used to calculate the extent of LD between these 17 variants and the GWAS SNPs located in *IDE* or its 100 kb flanking regions. This analysis showed that GWAS SNP rs7910977 was in strong linkage disequilibrium with rs6583817 ($r^2 = 0.996$ and $D' = 1$) identified in the current study. The measurements of *IDE* transcript levels described in the "Measurement of *IDE* mRNA expression" section above included 174 cerebellar samples obtained at autopsy from subjects that were part of the

LOAD GWAS. Consistent with the results reported for rs6583817, rs7910977 associated strongly with the level of *IDE* transcript in these 174 samples with a p value of 2.7×10^{-8} that was significant at the genome-wide level [375]. rs7910977 also showed significant association with LOAD in the 2,099 subjects analyzed in the GWAS (p = 0.005).

3.2.10 Association of rs7910977 with Plasma Aβ40 and Aβ42

rs7910977 was genotyped as part of two genome-wide association studies of the Dalmatian island populates of Vis and Korcula, using Illumina genotyping platforms [264, 271] Genotypes for this SNP were present (following quality control) for 883 of the 949 subjects with plasma Aβ measurements in the Vis study and 876 of the 930 subjects with plasma Aβ measurements in the Korcula study. Study samples from Vis and Korcula were collected as described previously (Chapter 2.) and plasma levels of the Aß40 and Aß42 peptides were measured, in duplicate, using a well established sandwich ELISA system, also described elsewhere (Chapter 2.). Antibody pairs BNT77/BC05 were used to measure plasma Aβ42 in both populations, the BAN50/BA27 pair was used to measure Aβ40 in the Vis population and the BNT77/BA27 pair was used to measure Aβ40 in the Korcula population. The ELISA assays were repeated for individuals that had a coefficient of variation (CV) of greater than 0.2 for the duplicate measurements; following repetition of these measurements the median value was used as the estimate of the plasma Aβ value. A third trait was created termed “total Aß” by summing these raw values for

Aß40 and Aß42 for each person. This trait was then analysed in the same manner as the Aß40 and Aß42 traits individually.

The R package GenABEL [318] was used for association analysis as described in chapter 2. The results of this analysis for rs7910977 is shown in table 3.3

3.2.11 Functional assessment of IDE variants

The effect of rs6583817 described in this study was evaluated for functional influence on gene expression levels using a dual luciferase reporter assay in Human HepG2 hepatocellular carcinoma and Be(2)-C neuroblastoma immortalized cell lines as described by Carrasquillo et al [3]

3.3 Results

3.3.1 Variant Discovery.

Using the variant discovery approaches outlined in 3.2.3 the Younkin group identified 17 common polymorphisms (minor allele frequencies >0.01) likely to have functional effects (**Figure 3.1**). These variants and the haplotypes they form were tested for association with *IDE* mRNA expression using transcript levels measured in 194 LOAD cerebella sampled at autopsy from the brain bank at Mayo Clinic, Jacksonville. They were also analyzed for genetic association with LOAD in a large LOAD case-control series with 3,529 AD cases and 3,441 controls drawn from four American and five North European series (data not shown).

3.3.2 Association of IDE SNPs and haplotypes with IDE transcript and LOAD.

The 17 *IDE* variants selected in 3.3.1 formed eleven haplotypes with frequencies >1% that could be tagged with ten variants. In the combined series, these haplotypes were significantly associated with *IDE* transcript levels (global $p = 0.003$) and with LOAD (global $p = 0.02$) See Carrasquillo et al [3].

Assessment of the IDE haplotype tagging SNPs for association with the level of IDE transcript identified nominally significant association with four of these SNPs. Association was most significant for rs6583817 and rs5786996 with p-values of 1.5×10^{-8} and 0.0005 respectively, and linear regression showed that for both of these variants the average level of *IDE* transcript was more than 2-fold higher for each minor allele carried (Carrasquillo et al). When these SNPs were later tested for association with LOAD, in the combined series, the minor alleles of rs6583817 ($p =$

0.03) and v3 (p = 0.02) were also found to be associated with reduced risk of LOAD. Notably, neither variant has previously been analyzed for association with LOAD, so they are not included in the AlzGene database. The association of these two variants with LOAD showed no series-related effects. Results for IDE transcript and LOAD association analysis for each of the 10 haplotype tagging SNPs is shown in Table 3.2.

3.3.3 Association of IDE variants with plasma Aß concentration

The involvement of *IDE* in LOAD pathogenesis was further explored by evaluating the effect of *IDE* on plasma Aß concentrations. This evaluation employed data from the studies fully described in Chapter 2. Plasma Aß40, Aß42 and total Aß (Aß40 +Aß42) were analyzed using 949 individuals from Vis and 930 individuals from Korcula (total = 1,879). The analysis took advantage of the strong linkage disequilibrium between rs6583817 and rs7910977 (r2 = 0.996 and D′ = 1), since rs7910977, but not rs6583817, was included in the genome-wide scan. It showed (**Table 3.3**) that rs7910977 was significantly associated with reduced levels of plasma Aß40 in the Vis population (ß ± SE = −0.147±0.06 p = 0.018) and suggestively associated with reduced Aβ40 in the Korcula population (ß ± SE = −0.086±0.08, p = 0.27). Meta-analysis of the two populations showed overall significant association with reduced Aß40 (ß ± SE = −0.124±0.05, p = 0.011). Analysis of the two populations (**Table 3.3**) also showed a suggestive association between rs7910977 and reduced plasma Aß42 levels (meta-analysis ß ± SE = −0.07±0.05, p = 0.18). Similarly, analysis of total Aß (Aß40 +Aß42) revealed

significant association with reduced levels in the Vis population (ß ± SE = −0.165±0.06, p = 0.011) and meta-analysis of the two populations showed overall significant association with reduced total Aß (ß ± SE = −0.130±0.05, p = 0.009). Thus the variant that showed strongest association with elevated *IDE* expression also showed significant association with reduced levels of Aß, a substrate of the insulin degrading enzyme that is implicated in LOAD pathogenesis.

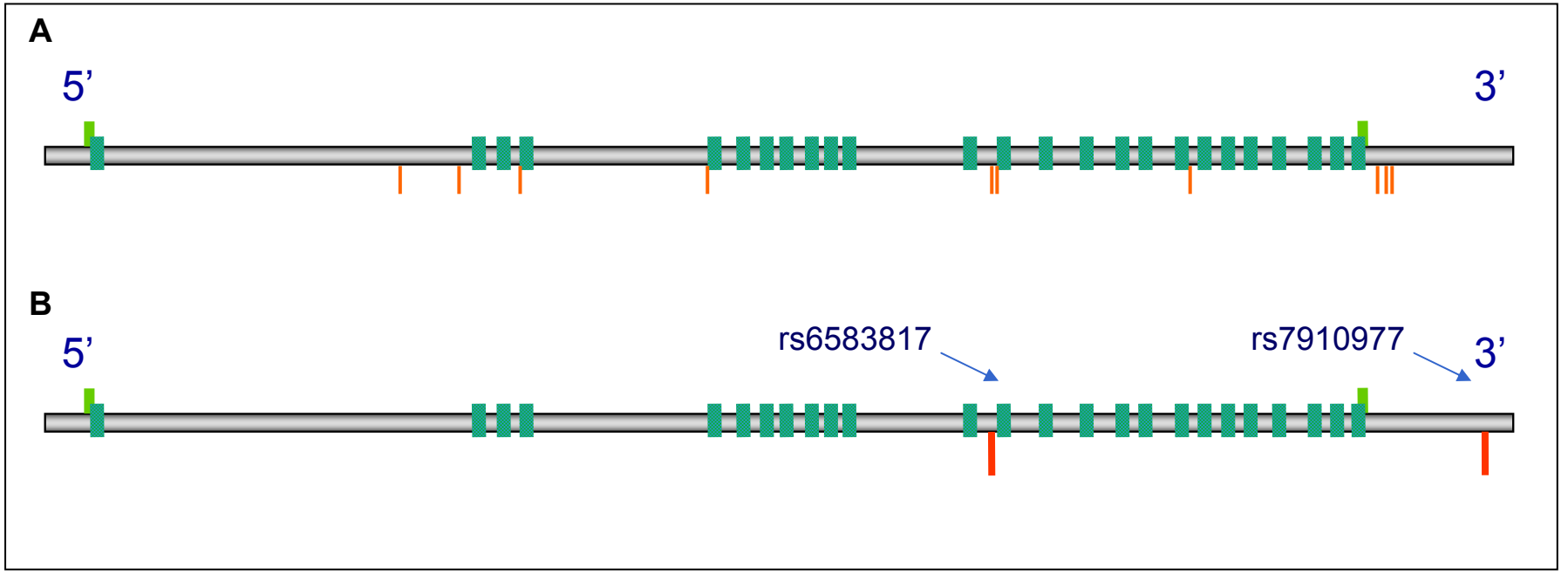

Figure 3.1 Schematic of the *IDE* gene. Exons are shown as blue bars, SNPs are shown as red bars, UTR regions are indicated as green bars. **(A)**Ten haplotype tagging SNPs outlined in section 3.3.1 are shown. (B) The location of two variants (rs6583817 & rs7910977) highlighted in this study are indicated.

						Association with IDE mRNA			Association with LOAD)			
rs number	**Position**	**Location**	**Cons**	**MAF**	**Minor Allele**	**β**	**Fold Δ (95% CI)**	**p**	**AD**	**CTRL**	**OR (95% CI)**	**p**
V2 (no rs#)	94,202,383	3′ flank	81%	0.02	del	−0.03	1.02 (0.60–1.74)	0.94	0	0.02	1.15 (0.88–1.50)	0.3
rs5786996	94,202,516	3′ flank	74%	0.04	C	−1.08	2.12 (1.40–3.21)	**0.0005**	0	0.04	0.77 (0.62–0.95)	**0.02**
rs5786997	94,203,071	3′ flank	88%	0.11	AT ins	−0.09	1.07 (0.78–1.45)	0.69	0.1	0.11	0.96 (0.84–1.09)	0.52
rs4646957	94,219,892	Intron18	71%	0.36	A	−0.28	1.22 (1.03–1.43)	**0.02**	0.4	0.36	0.98 (0.90–1.07)	0.67
V310 (no rs#)	94,236,972	Exon13	92%	0.08	T	−0.22	1.16 (0.68–1.98)	0.58	0	0.02	1.09 (0.80–1.48)	0.6
rs6583817	94,237,227	Intron12	74%	0.13	A	−1.08	2.12 (1.65–2.71)	**1.5×10^{-8}**	0.1	0.13	0.87 (0.78–0.99)	**0.03**
rs17875327	94,264,789	Intron4	85%	0.1	C	0.18	0.88 (0.67–1.15)	0.35	0.1	0.1	0.97 (0.84–1.10)	0.61
rs4646955	94,284,271	Intron3	76%	0.25	C	0.1	0.93 (0.78–1.12)	0.46	0.3	0.25	1.07 (0.98–1.17)	0.14
rs17107721	94,288,480	Intron1	71%	0.05	A	−0.07	1.05 (0.73–1.50)	0.8	0.1	0.05	0.88 (0.73–1.06)	0.18
rs11187061	94,295,389	Intron1	70%	0.18	T	0.35	0.78 (0.65–0.94)	**0.009**	0.2	0.18	1.07 (0.96–1.18)	0.2

Table 3.2 Association of IDE variants with cerebellar IDE mRNA and LOAD. Variant IDs and rs numbers (dbSNP variant identifier) are given for the 10 tagging variants; Chr position indicates relative position to the Human Genome build 36.1. Cons = conservation (>70% identity between the human and mouse sequence); MAF =minor allele frequency in controls. Single variant linear analyses vs. IDE mRNA levels and LOAD were performed in the same way that haplotypes were analyzed.

rs number (minor allele)	Study population	Aβ40			Aβ42			Total Aβ		
		beta	SE	p	beta	SE	p	beta	SE	p
rs7910977 (T)	Vis	-0.15	0.1	**0.02**	-0.1	0.1	0.1	-0.17	0.1	**0.01**
	Korcula	-0.09	0.1	0.27	-0.02	0.1	0.8	-0.08	0.1	0.31
	Meta Analysis	-0.12	0.1	**0.01**	-0.07	0.1	0.2	-0.13	0.1	**0.01**

Table 3.3. Association of rs7910977 with decreased plasma Aβ40 and Aβ42. Analyses were performed as described in Chapter 2.

		rs6583817 / rs7910977	
Phenotype	**N**	**P-value**	**Direction of association**
IDE expression (Cerebellum)	194	1.50E-08	Increase
In vitro reporter assays	2 cell lines	0.006	Increase
Plasma Aβ40	1879	0.01	Decrease
LOAD	3,529 AD: 3,441 CON	0.03	Decrease

Table 3.4. Summary of the findings presented in chapter 3 adapted from Carrasquillo et al [3]. This result illustrates how a quantitative phenotype, e.g. IDE mRNA levels, can show a far stronger association than the disease phenotype, e.g. LOAD. Given the evidence presented here we conclude that *IDE* plays a role in LOAD, at least in part through an Aß degradation pathway, although the effect size is small and thus difficult to replicate in LOAD case control studies.

3.4 Summary

In this study, conserved regions of IDE and its 10 kb flanks were examined in 269 AD cases and 252 were controls, thereby identifying 17 putative functional polymorphisms. These variants formed eleven haplotypes that were tagged with ten variants. Four of these showed significant association with *IDE* transcript levels in samples from 194 LOAD cerebella. The strongest, rs6583817, which has not previously been reported, showed unequivocal association (p = 1.56E-08, fold-increase = 2.12,); the eleven haplotypes were also significantly associated with transcript levels (global p = 0.003, data not shown). Furthermore, using data from the genome-wide studies described in Chapter 2, a proxy (rs7910977) for rs6583817 was found to associate significantly with decreased plasma Aß40 levels (ß =-0.12, p = 0.01) and total measured plasma Aß levels (β =0.13, p = 0.01). Finally, rs6583817 was associated with decreased risk of LOAD in 3,529 AD cases and 3,441 controls. (OR = 0.87, p = 0.03), and eleven IDE haplotypes (global p = 0.02) also showed significant association. Thus, a previously unreported variant unequivocally associated with increased IDE expression was also associated with reduced plasma Aß40 and decreased LOAD susceptibility. Genetic association between LOAD and IDE has been difficult to replicate. These findings suggest that targeted testing of expression SNPs (eSNPs) strongly associated with altered transcript levels in autopsy brain samples may be a powerful way to identify genetic associations with plasma Aß and with LOAD that would otherwise be difficult to detect. The implications of this study are more fully discussed in Chapter 5.

Chapter 4.

4.1. Background

Increasing age is a major risk factor for development of Alzheimer's disease. The mitochondrial theory of aging highlights the role of mitochondria in the aging process [275, 276, 382], making the study of mitochondria and mitochondrial dysfunction in age related diseases highly relevant. Furthermore, oxidative stress is widely believed to play a role in the aetiology of LOAD which is tightly related to mitochondrial function given that oxidative phosphorylation (OXPHOS) is a major source of reactive oxygen species (ROS) in the cell. The evidence for a mitochondrial role in LOAD was introduced in Chapter 1. As previously mentioned, the mitochondrion has its own genome with 37 genes which encode 13 proteins essential for the electron transport chain, 2 ribosomal RNA and 22 tRNA genes. The remaining proteins essential for the function of the electron transport chain are produced by genes in the nuclear genome and are post-translationally transported to the mitochondria. Additionally, many genes that produce antioxidant enzymes and proteins essential for mitochondrial replication and mtDNA integrity are located in the nuclear genome; consequently both the mitochondrial and nuclear genomes play an important role in mitochondrial function and mtDNA integrity. Overall there are more than 1000 proteins in the human mitochondrial proteome [383]

The aim of this chapter is to explore approaches for studying mitochondrial genetics in Alzheimer's disease and test these in pilot studies, in order to establish methods that can be applied to larger additional studies. This chapter covers three approaches

to evaluating the genetics of mitochondria with respect to aging and Alzheimer's disease pathogenesis:

i) An evaluation of mitochondrial somatic mutational burden as it relates to aging; adaptation of a method for measuring this which can be applied to DNA isolated from various tissues, such as brain, in future studies.

ii) Investigation of methods for measuring mitochondrial DNA damage and development of an adapted method; applying this to the study of mitochondrial DNA damage in samples of pathologically confirmed normal and diseased brain tissue.

iii) A candidate gene study is described, using an existing LOAD GWAS [3] to evaluate SNPs in known genes that are essential for optimal functioning of the mitochondrial electron transport chain.

Somatic mutational burden: In each cell there exists two copies of each of the 22 nuclear chromosomes, plus 2 sex chromosomes; however many mitochondria are present in each cell and within each there exists many copies of the mitochondrial genome, consequently the number of copies of the mitochondrial genome within each cell is both variable and greater than the nuclear genome. It has been hypothesized that mitochondrial DNA may act as a redox sensor for the cell. Accumulation of somatic mutations, over time, due to damage from ROS, that individually result in subtle changes to mitochondrial function, may collectively constitute a mutational burden, and result in the triggering of apoptosis [384].

Consequently, the study of somatic mitochondrial mutational burden is of interest to the aging and age-related disease fields. Due to the high copy number of the mitochondrial genome within each cell, traditional sequencing methods would fail to identify rare somatic mutations present in only a small number of individual genomes and so to address this issue several groups have made use of variations of a PCR-cloning-sequencing method [287, 385-387]. This method involves PCR amplification of a target locus in the mitochondrial genome, followed by cloning of the PCR products and selection of individual colonies for sequencing, thereby allowing the sequencing of PCR products from individual mitochondrial genomes. This approach makes it possible to identify variation at the level of a single mitochondrial genome from the many in a sample that are isolated from either a single cell or tissue.

The PCR-cloning-sequencing approach was tested using DNA isolated from blood samples of 13 control subjects ranging in age from 0 years (umbilical cord blood) to 97 years to investigate both the need for a high fidelity PCR polymerase and the link between mitochondrial mutational burden and aging, in addition to comparing the mutational burden at two different mitochondrial loci, in readily available DNA samples.

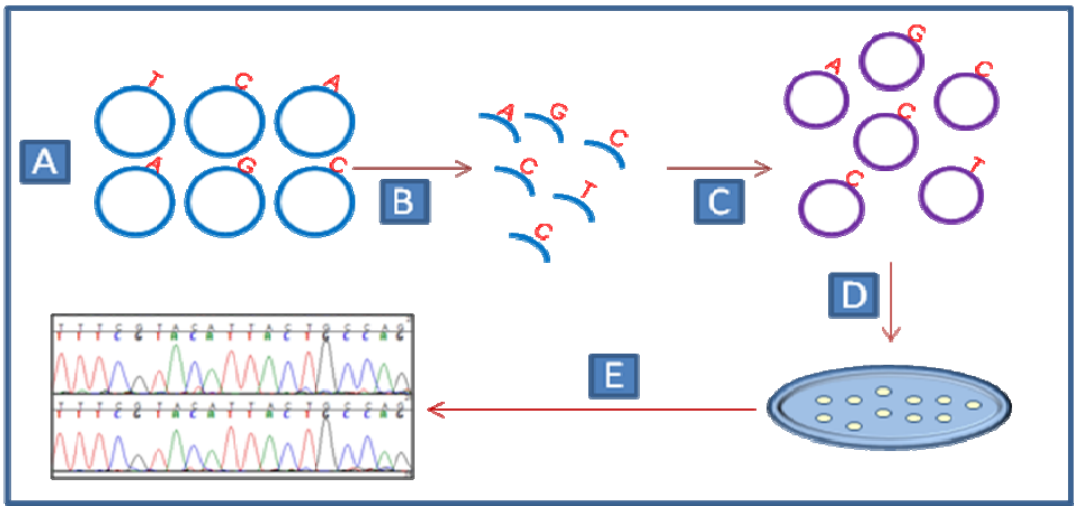

Figure 4.1. An illustration of the PCR-Cloning-Sequencing method for evaluating mitochondrial mutational burden. **A)** Six mitochondrial genomes are represented with various somatic mutations illustrated by the letters A,G,C and T. **B)** PCR amplification of this locus generates PCR products with a representative proportion of each mutation. **C)** Ligation of these PCR products into a suitable vector generates individual clones with a representative proportion of each mutation. **D)** Transformation and growth of competent cells generates colonies of cells with multiple copies of a single PCR product generated from a single mitochondrial genome. **E)** Sequencing of the isolated vectors from individual colonies facilitates the identification of mutations isolated from a single mitochondrial genome.

Mitochondrial DNA damage: In an extension of the somatic mutational burden hypothesis for mitochondrial dysfunction in aging and disease, researchers are investigating the role of mitochondrial DNA damage. In these studies, molecular lesions related to oxidative damage are the primary focus with 8-oxo-G being the most commonly measured lesion. However a more general approach to measure DNA damage has also been developed that involves the use of quantitative PCR (qPCR) where the premise is that molecular lesions, resulting from oxidative damage, will reduce the efficiency of the PCR reaction in a way that is measurable. This approach has been reported successfully by others previously [386] and has been tested and adapted here for measuring mtDNA damage from DNA isolated

from brain tissue. Dr. Dennis Dickson at the Mayo clinic in Jacksonville has a large collection of Autopsy confirmed AD and non-AD brains. Tissue was collected from 40 elderly subjects with one of three different disease pathologies (AD, PSP or VaD) or no disease pathology. The pathological changes that occur in LOAD are more severe in the temporal cortex than in the cerebellum, so DNA was isolated from tissue from both regions of the brain to determine if any observations are localized to affected tissue in LOAD or are more generalized. The comparison with other neurodegenerative diseases, in addition to normal elderly controls, can determine if any observations are specific to AD or a more general feature of neurodegenerative disease.

In previous publications, the extent of mtDNA damage has been determined by measuring a nuclear locus (nDNA) and one or more mitochondrial loci and determining the damage by calculating the ratio of mtDNA to nDNA [386]. This approach has been adapted here to allow for differences in the concentration of nuclear DNA between samples without this affecting the calculation of mtDNA damage. The total [DNA] was measured for each sample using spectrophotometric methods and the nuclear contribution to this value was quantified using nuclear targeted qPCR assays. An expected concentration of mitochondrial DNA is generated by subtracting the quantified nuclear concentration from the concentration as determined spectrophotometrically. The damage at each mitochondrial locus of interest is then defined as the difference between the expected amount of mtDNA in the sample (total minus nDNA) and the observed amount of mtDNA in the sample (measured using qPCR). This difference is presented as a proportion of the expected

amount of mtDNA and as such represents the proportion of damage. Like the other approaches using qPCR methods, this approach does assume that the nuclear locus does not exhibit significant DNA damage.

Additional considerations for mitochondrial genetic studies: When investigating mitochondrial DNA there are various considerations that need to be taken into account to reduce the effect of experimental artefacts on the results of such studies. Primarily, primers targeting mitochondrial loci need to be designed carefully and tested where possible to prevent the amplification of nuclear DNA (numts) or nuclear pseudogenes. In the studies described in this chapter the oligo's are tested for mitochondrial specificity by including a ρ^0 DNA sample in all experiments. The ρ^0 sample was isolated from a human osteosarcoma 143b cell line that hade been treated to remove all mitochondrial DNA [388]. The generation of PCR product from this DNA could only result from nuclear DNA, providing a test for the specificity of pairs of primers directed to the mitochondrial genome.

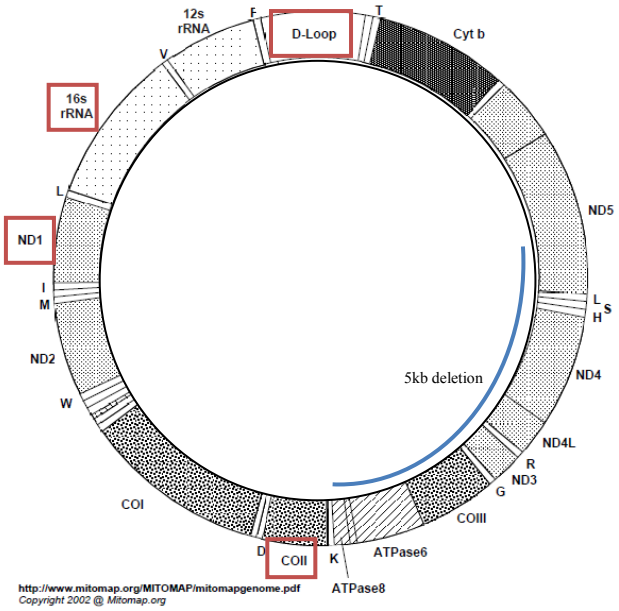

Figure 4.2. Mitochondrial genome. Figure adapted from MITOMAP [1]: mitochondrial loci investigated in this chapter are highlighted in red boxes. The common (Kearns Sayre syndrome, KSS) 5kb deletion is shown: none of the genes selected in this study overlap with this deletion reducing the risk of the presence of this deletion being detected in the experimental samples and confounding the results.

Careful consideration of DNA isolation methods is needed to prevent oxidative DNA damage during the isolation process. Following tests of various DNA isolation methods Helbock et al [389] provide evidence in support of the sodium iodide method to prevent *in vitro* DNA damage. Consequently, for the mitochondrial DNA damage study, DNA was isolated from brain tissue using a Kit available from Wako Chemicals Inc. This protocol includes the preparation of a crude mitochondrial sample, using differential centrifugation, prior to extraction of DNA from the sample and the use of sodium iodide for DNA isolation. The advantages of using this approach include having a more concentrated mtDNA sample, with expected lower levels of nuclear DNA in the sample and reduced *in vitro* DNA damage.

The final consideration made in the design of the studies described here was the evaluation of the need for a high fidelity PCR polymerase in the initial PCR step of

the PCR-cloning-sequencing approach used for studying mitochondrial mutational burden. As can be seen from the results of this study described later, the PCR polymerases that do not have proof-reading capabilities introduce variation during the PCR process that is detected in the sequencing step, so that high fidelity polymerases are necessary to improve the reliability of the results for these types of studies.

Nuclear candidate gene association study: A review of the current literature and the Alzgene database of Alzheimer's disease studies, indicates a lack of genetic association studies evaluating candidate genes involved in mitochondrial function and mtDNA integrity. Since a major focus of this chapter is on mitochondrial DNA damage and mutation, a candidate gene study of the nuclear genes that encode the mitochondrial DNA polymerase is relevant and SNPs in three genes involved in mitochondrial DNA replication (*POLG*, *POLG2* and *PEO1*) are studied here for association with LOAD. Furthermore, there is a wealth of literature reporting deficiencies in the mitochondrial electron transport chain in Alzheimer's patients, especially for the cytochrome c oxidase (or complex IV) enzyme [302, 303, 390, 391]. The three catalytic subunits of this complex are encoded by the mitochondrial genome (COX I - COX III), but the remaining 10 structural subunits are encoded by genes in the nuclear genome and are excellent candidates for a LOAD association study. A literature search did not identify any previously published reports of SNP based association studies with these genes and LOAD, so this study would present a novel candidate gene study in LOAD. Chandrasekaran.K et al [392] evaluated the expression of both mitochondrial and nuclear genes encoding subunits of the

OXPHOS pathway and reported changes in expression of COX I, III (mitochondrial) and IV (nuclear) in different regions of the brain in AD versus control subjects, providing further evidence for a role of both nuclear and mitochondrial genes encoding subunits of complex IV of the ETC in LOAD. In this chapter, 17 nuclear genes that encode the 10 nuclear subunits of complex IV are included in an association study, making use of data from a LOAD GWAS published in 2009 [3] that includes 788 AD subjects and 1175 control subjects aged between 60 and 80 years of age. The outcome of this study highlights SNPs for follow-up analysis in additional LOAD case control series.

Additional approaches: In addition to the approaches tested in this chapter, there have been other methods developed for the study of mitochondrial DNA. It was beyond the scope of this chapter to test and evaluate all of these approaches for studying mitochondrial genetics but the samples prepared in this study can be used for additional studies such as next generation sequencing of the mitochondrial genome with platforms such as the MitoChip (Affymetrix) or ABI Solid (Applied Biosystems).

4.2. Methods

4.2.1. Mutational burden study.

Study samples: Initially, DNA samples isolated from the white blood cells of three control subjects, aged 0 (umbilical cord blood), 59 years and 97 years, that were supplied by Dr. Caroline Hayward, were collected for a pilot study of the PCR-cloning-sequencing method (Figure 4.1.). An additional 10 control subjects were later selected to further evaluate the relationship between age and mitochondrial mutational burden in blood as a test model. All thirteen subjects and the age groups in which they were categorized for analysis are detailed in Table 4.1. The DNA for all 13 control subjects described in this study were isolated from blood by others. All subjects included in this study gave informed consent.

Pre-cloning PCR: PCR primers were designed flanking the *HVR* and *ND1* regions of the mitochondrial genome; generating PCR products of 999 bps and 1157 bps respectively. The primers are detailed in the appendix (Table i). Three polymerases were tested for each primer pair: Phusion (Finnzymes, Finland: part# F-540S), Platinum (Invitrogen, UK: part# 10966-034) and AmpliTaq Gold (Applied biosystems, UK, part# N8080246) According to the manufacturer Phusion has an estimated error rate of 4.4 $x10^{-7}$ nucleotides (nt) in the buffer provided. The Platinum and AmpliTaq Gold polymerase manufacturers do not estimate the error rates, but published error rates of non-proof reading Taq polymerase enzymes have been estimated to be between 2.4 $x10^{-4}$ and $2.0x10^{-5}$ nt [393, 394]. A control DNA sample isolated from cells lacking mitochondria (osteosarcoma 143b ρ^0 cell line) was also tested with each set of primers in order to evaluate them for mtDNA

specificity by checking for the presence of a PCR product on a 2% TBE gel. The presence of PCR product using the ρ^0 template would indicate the amplification of nuclear pseudogenes. PCR cocktails were prepared using the buffers supplied with each of the respective polymerases, and 10mM dNTPs (Qiagen, USA, Cat# 201901) was added to the Platinum and Phusion PCR cocktails to a final concentration of 200µM. The Amplitaq Gold reaction made use of a PCR mastermix which did not require the addition of dNTPs. The relevant primers were included in each of the three PCR reactions at a concentration of 10µM, with a final concentration of 0.5µM in each reaction. Thermocycling conditions for the Phusion polymerase PCR reactions were 30 seconds at 98^{o}C, followed by 30 cycles of 10 seconds at 98^{o}C, 30 seconds at 58^{o}C and 30 seconds at 72^{o}C, with a final extension of 72^{o}C for 8 minutes. The thermocycling conditions used for the Platinum and Amplitaq Gold PCR reactions were 6 minutes at 94^{o}C, followed by 30 cycles of 30 seconds at 94^{o}C, 30 seconds at 55^{o}C and 60 seconds at 72^{o}C, with a final extension of 72^{o}C for 10 minutes.

Cloning: In order to evaluate individual mtDNA molecules for somatic mutations a cloning approach was used, similar to that used by others previously [287, 385-387]. PCR products generated using Phusion polymerase were incubated with a non-proof reading Taq polymerase (Qiagen, USA, part 201203) at 72^{o}C for 10 mins in order to add TA ends for cloning into the pGemTeasy vector (Promega, UK, Catalog# A1360). PCR products were purified prior to being cloned into the pGemTeasy vector using the Qiaquick PCR purification kit spin protocol (Qiagen, USA, part#28104), according to the manufacturer's directions. Ligation of PCR products

was carried out according to the manufacturer's directions with some modifications with respect to the PCR product-to-vector ratio which was optimized for each reaction. Competent cells (Top10 competent cells, Invitrogen, USA, cat# C4040-03) were transformed for each separate sample, according to the manufacturer's instructions, and 200µl, 100µl or 20µl of the transformed cells (with added SOC) were plated on three agar plates with 100µg/ml ampicillin (pre-treated with 20µl X-Gal for blue/white selection), then grown at 37^{o}C overnight in an incubator.

Post-cloning-PCR: Following bacterial cell growth, 95 individual colonies were picked for each of the three DNA samples and types of polymerase (a total of nine 96 well plates), using sterile techniques, into 15ul sterile water previously aliquoted into the wells of a 96 well PCR plate (Greiner bio-one, UK, Cat# 651550). The selected colonies were left at room temperature for 1 hour to facilitate cell lysis and then used as the template for a PCR using M13 primers, an approach used previously by others [387]. The PCR was carried out using a standard PCR reaction kit (Qiagen, USA, Cat# 201205) including 10mM dNTPS (Qiagen, USA, Cat# 201901), after which 5ul of PCR cocktail was added to 15ul of dsH_2O into which the colonies had been picked. The thermocycling conditions were then as follows: 2 minutes at 94^{o}C, 35 cycles of 30 seconds at 94^{o}C, 30 seconds at 55^{o}C and 90 seconds at 72^{o}C, with a final extension of 72^{o}C for 5 minutes.

Sequencing: PCR products, generated using M13 PCR primers, were purified using filter plates, according to the manufacturer's directions (Millipore, USA, Cat# MSNU03010) on a MultiScreen™ Vacuum Manifold 96-well (Millipore, USA, Cat# MAVM0960R). The sequencing reaction was prepared using Applied

Biosystems BigDye® Terminator v3.1 Cycle Sequencing kits, slightly differently from the manufacturer's directions, based upon data from experiments carried out by Stuart McKay at the MRC Human Genetics Unit (data unpublished). The sequencing reaction cocktail was prepared as follows; for each reaction, 0.25ul of Big Dye Terminator, 1.86ul 5x buffer, 1ul primer (3.2pmol) and 4.89ul dsH_2O was added to 2ul purified PCR product. Otherwise the protocol was followed as directed by the manufacturer. The cycle sequencing was carried out using 96 well PCR plates with the following cycle sequencing protocol: 94^oC for 10 seconds, 50^oC for 5 seconds, and 60^oC for 4 mins, repeated 25 times followed by a 12^oC hold. The cycle sequencing products were purified using ethanol precipitation, prior to being sequenced using a 3100 ABI genetic analyzer in both forward and reverse directions using M13 sequencing primers. The ethanol precipitation was carried out according to the directions supplied by Applied Biosystems 3100 genetic analyzer manual (chapter 3 page 11). Precipitated PCR products were re-suspended in 10ul Hi-Di formamide (Applied Biosystems, Cat# 4311320) and transferred to 384 well plates for sequencing (Applied Biosystems, USA, Cat# 4309849). Analysis of sequencing data was carried out using the software program Sequencher (Gene Codes, USA).

Quality control of sequencing analysis: Clones that failed sequencing in one direction were removed from analysis in the opposite direction; mutations that were detected within the first 25 base pairs of sequence, including the primer sequences of HVR and ND1 were not included in the analysis, as frequently the early sequence had high background ("noise") levels. Contigs were constructed within the Sequencher program with sequenced samples aligned against the reference mtDNA

sequence. The total number of base pairs analyzed was determined by the point at which sequence could not be reliably read within a contig; any overlap between forward and reverse contigs was subtracted from the total base pair count. All mutations were counted once (if they appeared in the overlap of sequence between the forward and reverse directions of sequence then it was noted that they were confirmed in both directions but only counted once) and the number of clones in which they appeared was noted.

Additional samples: To further investigate the relationship between age and mitochondrial mutational burden in lymphocytes, the PCR-cloning-PCR-sequencing approach, described in the previous pages, was used for an additional 10 samples to investigate the HVR locus only. These samples are described in Table 4.1. The same PCR primers were used for the HVR locus as were used previously, and the PCR protocol for only the Phusion polymerase (which has proof reading capabilities) was used. Since only the Phusion polymerase was used, for the first PCR step, a Zero Blunt® PCR cloning kit (Invitrogen, USA, Part# K2700-40) and One Shot®Top10 competent cells (Invitrogen, USA, cat# C4040-03) were used, instead of adding TA ends and cloning in the pGEM-T-easy vector as before. Transformed cells were plated onto agar plates with 50ug/ml kanomycin (100ug/ml ampicillin plates were used previously). In order to facilitate the sequencing of a larger number of samples, 47 clones for each locus and subject were selected, (95 clones were selected previously). The PCR, cloning and sequencing procedures were identical to those described previously except where noted above.

Statistical analysis: For each subject, the mitochondrial mutational burden was calculated as the number of heteroplasmic mutations identified in 2 or fewer clones per subject, divided by the total number of nucleotides evaluated for that subject, multiplied by 1.0E06. To evaluate the results of the pilot study, the Kruskal-Wallis test was implemented to test the effect of the polymerase across all three samples at each of the HVR and ND1 loci; a significant p-value with this test would indicate a rejection of the null hypothesis that the polymerase has no effect on the detected somatic mutational burden. The Kruskal-Wallis test was also used to evaluate the variance of the mutational burden at the HVR locus across the four age groups, in the second study.

4.2.2. mtDNA damage study.

Sample selection: In Alzheimer's disease, the cerebellum is not usually affected by AD pathology, whereas the temporal cortex is frequently affected by AD pathology, for this reason, this study focused on brain homogenates generated from frozen brain tissue, sampled from the cerebellum and temporal cortex. Tissue was provided by Dr. Dennis Dickson, Mayo Clinic, Jacksonville, FL. Subjects were chosen from over 1,500 subjects for whom frozen brain tissue was available from the temporal cortex and cerebellum. Study subjects without mixed pathology were preferred and so sample selection was limited: there were 15 elderly normal control subjects and 175 elderly AD subjects that fitted this criterion.

Sample #	Age (years)#	Sex	Group
1	43	Male	
2	46	Female	A
3	46	Female	
4*	59	Male	
5	60	Male	B
6	60	Female	
7	80	Male	
8	80	Male	C
9	80	Female	
10	92	Male	
11	93	Female	D
12*	97	Female	
13*	0	Female	NA

Table 4.1. **Samples and groups used to investigate the role of age on mitochondrial mutational burden**. # Age indicates the age of the subject at which the blood sample was drawn. * Indicates the three samples used for the pilot study.

Among subjects with non-AD pathologies, again without mixed pathology, the most frequent were progressive supranuclear palsy (PSP), with 91 elderly subjects and vascular dementia (VaD), with 25 elderly subjects. Consequently, PSP and VaD subjects were selected as the non-AD dementias in this study. All 40 subjects selected for this study are detailed in Table 4.2. All AD's had a Braak stage of 6 and all other subjects had a Braak stage between 0 and 3 [12].

DNA isolation: In a 1998 Helbock et al. [389] described the effects of DNA isolation methods on DNA damage, indicating that phenol:chloroform based methods can cause *in vitro* DNA damage and concluding that a DNA isolation method that makes use of sodium iodide causes the least *in vitro* DNA damage during DNA isolation. For this reason, the Wako mtDNA Extractor CT Kit (Wako Chemicals Inc., USA, Catalog# 291-55301) was chosen to isolate a crude mitochondrial DNA preparation from the frozen brain tissue samples. This kit facilitates the isolation of mitochondria-enriched DNA samples using the sodium iodide approach so it should reduce *in vitro* DNA damage when compared with alternative methods.

Sample#	Path Dx	Age (years)	Sex
1	**AD**	76	M
2	**AD**	78	M
3	**AD**	83	M
4	**AD**	86	F
5	**AD**	87	F
6	**AD**	88	F
7	**AD**	88	F
8	**AD**	90	M
9	**AD**	90	F
10	**AD**	97	F
Mean		86.3	

Sample#	Path Dx	Age (years)	Sex
11	**Normal**	74	M
12	**Normal**	76	M
13	**Normal**	78	F
14	**Normal**	83	M
15	**Normal**	86	F
16	**Normal**	90	F
17	**Normal**	92	M
18	**Normal**	93	F
19	**Normal**	93	F
20	**Normal**	97	F
Mean		86.2	

Sample#	Path Dx	Age (years)	Sex
21	**PSP**	78	M
22	**PSP**	78	M
23	**PSP**	79	F
24	**PSP**	79	F
25	**PSP**	81	F
26	**PSP**	81	M
27	**PSP**	82	F
28	**PSP**	83	F
29	**PSP**	91	F
30	**PSP**	98	F
Mean		83	

Sample#	Path Dx	Age (years)	Sex
31	**VaD**	74	F
32	**VaD**	76	F
33	**VaD**	83	M
34	**VaD**	84	M
35	**VaD**	87	F
36	**VaD**	87	F
37	**VaD**	89	M
38	**VaD**	90	F
39	**VaD**	90	F
40	**VaD**	91	F
Mean		85.1	

Table 4.2. Subjects chosen for study of mtDNA damage. AD and control subjects were matched, based on age at death. Subjects were also matched on *ApoE* genotype within the constraints of the limited number of available subjects. AD: Alzheimer's disease, Normal: No pathology that meets disease phenotype criterion, PSP: progressive supranuclear palsy, VaD: vascular dementia. M= Male, F=Female. Age is the age at death.

mtDNA was isolated from frozen brain tissue according to the manufacturer's protocol. The concentration of each DNA sample was determined in triplicate measurements using the NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) in order to get a reliable DNA concentration measurement.

Selection of loci for qPCR measurement: Applied Biosystems has pre-developed assays for quantifying nuclear DNA concentration for two single copy genes: *RNAseP* and *β-actin*. To measure mitochondrial DNA concentration, three loci were chosen: *16s, D-loop* and *COII*. Loci within the common KSS 5kb deletion were not selected for this study to avoid confounding arising from the presence of this deletion. The *16s* locus was a neutral locus with no *a priori* reason for involvement in LOAD. The *D-loop* locus was of interest due to the fact that it is involved in mtDNA replication, while the *COII* gene was of interest due to the reported association between complex IV deficiencies and LOAD.

Cloning of D-loop, COII and 16s loci for standard curve: In order to perform quantitative PCR, a known amount of the target mitochondrial loci (*D-loop, COII* and *16s*), was needed as a control for generating a standard curve. This was achieved by cloning the target loci from a cord blood DNA sample based upon the assumption that the cord blood DNA would exhibit the least baseline DNA damage when compared with older samples. 10ng of DNA isolated from cord blood (Table 4.1) and was PCR amplified using primers for the *16s and D-loop* and *COII* loci listed in Table i (appendix). The Phusion proof-reading polymerase and its supplied buffer was used for PCR amplification, as described previously. A negative (no DNA template) control and ρ^0 DNA sample were also included to check for PCR

contamination and amplification of nuclear pseudogenes respectively. PCR products were visualized and separated by running 50µl of PCR product on a 2% TBE gel stained with ethidium bromide. The target PCR products were purified from the gel using the Qiaquick gel extraction kit (Qiagen, USA, Cat.no 28704) according to the manufacturer's protocols. Purified PCR products were cloned using the Zero Blunt® PCR cloning kit (Invitrogen, USA, Part# K2700-40) and One Shot®Top10 competent cells (Invitrogen, USA, cat# C4040-03). These cells were transformed with the resultant plasmids, plated on agar plates containing kanomycin (50ug/ml) and incubated at 37^{o}C overnight. Five colonies were selected and transferred into five vials of Luria-Bertani (LB) medium containing 50ug/ml kanomycin and grown overnight in a shaking incubator at 37 ^{o}C. The plasmids were then purified using a QIAprep Spin Miniprep kit (Qiagen, USA, cat.no 27104) according to the manufacturer's instructions. The purified plasmids were sequenced, using M13 sequencing primers (Table i), to verify the presence of the correct sequence insert. Sequencing was carried out by the sequencing core facility at the Mayo Clinic in Rochester Minnesota, USA.

Quantitative PCR of 16s D-loop and CoxII loci to measure DNA damage: Applied Biosystems Taqman assays were designed to quantify each of the three mitochondrial loci chosen for investigation of mtDNA damage. The primers and probes are listed in Table i. Oligonucleotides for qPCR of the *D-loop* and *COII* loci were selected from a previous publication by Mambo.E et al. [386], which had previously been successful in quantifying these loci. In addition to mitochondrial target assays, absolute quantification of the *RNAseP* nuclear locus (Applied

Biosystems cat.no 4316831) and the *β-actin* nuclear locus (cat. no 401846), was carried out using pre-designed Taqman assays obtained from Applied Biosystems. 384 well PCR plates (Applied Biosystems, USA, part# 4309849) were used to facilitate the measurement of all samples, standard curve dilutions and negative and ρ^0 DNA controls, each in triplicate, using one plate per assay. Only the centre wells on the plate were used in order to avoid the possibility of evaporation affecting the results. PCR cocktails were prepared using 5µl of Taqman Universal PCR master mix (Applied Biosystems, USA, cat. no 4364338) in addition to amplification primers (forward and reverse, 900nM each), probe (250nM) and dsH2O up to a volume of 9µl per reaction, which was added to 1µl of DNA or dsH20 (negative controls) plated in each well, so that each reaction had a volume of 10µl. For mitochondrial assays, 2ng of DNA was plated in each well and for the *RNaseP* and *β-actin* assays, 10ng of DNA was plated in each well. One µl of standard curve dilutions, at the relevant concentrations, were also included on the plate, as described below.

Standard curves: For the RNAseP and β-actin assays, a cord blood DNA sample, which had been used as the template for cloning of the mitochondrial targets, was used to generate a standard curve ranging from 20ng to 0.02ng using a 1:2 serial dilution, generating 11 data points on the standard curve, plus an additional point with inclusion of the zero template controls. For the mitochondrial assays of *16s, D-loop* and *COII* loci, the standard curve was based upon the number of target molecules generated by the cloned genes in the respective plasmids. The number of copies of each target gene can be calculated based upon the ng of DNA purified

from the mini-prep for each clone and the known size (in nucleotide base pairs, bps) of each plasmid with the target insert. The sample was then diluted to generate a standard curve of copy numbers of each gene. Since 2ng of DNA was plated for each of the study samples and a maximum contribution of the DNA in the sample is 100% mitochondrial DNA, assuming each genome is 16kb, the maximum number of molecules that can be in 2ng of mtDNA is 1.14E-08 molecules. Consequently the upper value of the standard curve was set to 1E-10 molecules and a standard curve using a serial dilution of 1:10 was generated, down to an estimated single molecule which provided 11 data points on the standard curve. This was done separately for each cloned locus (*16s*, *COII* and *D-loop*) and only the relevant clone was present on the plate as a standard curve for the target assay.

Assessment of numts using QPCR: For each of the three mitochondrial absolute quantification assays, ρ^0 DNA was included on each plate in triplicate (2ng/well). The number of amplified copies of nuclear mitochondrial DNA pseudogenes (numts), was measured and quantified using the same approach as for the experimental samples. The amplification of numts was compared to the background levels, as measured by the negative control, and also in relation to the minimum value measured for the experimental samples.

Data analysis: Applied Biosystems SDS software was used for the initial data analysis prior to exporting the data to Microsoft Excel (Seattle, WA) for further analysis. The SDS software facilitates the transfer of Ct values (raw data) measured for each well on the plate into an estimate of the number of molecules in each well using the standard curve. These values are exported to Excel and the median value

for each triplicate measure for each sample was determined and subsequently used as the estimate of the number of molecules measured (in the case of the mitochondrial assays) and of the weight in ng (for the nuclear assays).

The amount of nuclear DNA (ng) for each sample was determined by the mean of the triplicate values (ng) measured by the RNAseP and β-actin assays. The proportion of DNA for each sample that can be accounted for by nDNA was calculated by dividing the average measured value by 20 (20ng of DNA were plated in each well for the nuclear assays). This proportion was then used to determine the amount of mtDNA that was expected to be measured by the mitochondrial absolute quantification assay.

The amount of measured mtDNA (ng) was determined for each locus individually by dividing the number of measured molecules by 5.51E+07: this value is the number of full length mtDNA molecules per ng of DNA and was determined as follows:

- The number of base pairs per full length mitochondrial genome molecule is 16,568, which is equal to a molecular weight of 10.9 million Daltons.

- A 1E+06 Dalton molecule is equal to 1ug/pmole, so one mtDNA molecule is equal to 10.9 ug/pmole.

- 1 pmole has 6.02E+11 molecules: consequently 10.9ug of mtDNA is equal to 6.02E+11 molecules and 1ng of mtDNA is equal to 5.51E+07 molecules.

The level of damage for each mitochondrial locus and sample was then determined as the difference between the expected amount of mtDNA and the observed amount

of mtDNA, divided by the expected amount of mtDNA (translated to a percentage by multiplying this value by 100). For example: If the average amount of nDNA that was measured for sample A containing 20 ng total DNA was 5ng, then it is expected that 15ng (or 75%) of the DNA in this sample is mtDNA. Therefore, when 2 ng of DNA is plated in a well, it is expected that 1.5ng of this is mtDNA. If the assay at the *16s* locus measures 0.5ng of mtDNA, then the difference between what was expected and what was observed is 1ng, which is equal to a 67% reduction in mtDNA compared with the expectation.

In order to determine if there was variation in the measured damage at each of the three mitochondrial loci, Microsoft Excel was used to calculate the correlation co-efficient (r) between the damage values at each of the three pairs of loci for each set of 10 samples categorized by tissue and disease diagnosis.

To examine if any of the three mitochondrial loci tended to have more DNA damage than others, Friedman's test was implemented, using StatsDirect software (StatsDirect Ltd, England) for each set of 10 samples, categorized by tissue and disease.

In order to determine if any of the groups exhibited a significantly different level of DNA damage from that of the 'normal' samples, the ratio between the mean DNA damage value for each group of 10 samples with a neurodegenerative disease (AD, PSP and VaD) and the mean DNA damage value for the 10 samples with no neurodegenerative disease (Normal) was calculated and the Mann-Whitney (Wilcoxon rank sum) test was implemented in StatsDirect to determine the significance of the difference between these mean values.

Likewise, to evaluate the relative levels of DNA damage between the cerebellum and temporal cortex of each set of 10 samples at each mitochondrial locus, the ratio between the mean values were calculated and the Wilcoxon signed rank test was implemented in StatsDirect to determine the significance of these differences.

4.2.3. Candidate gene study

Subjects: The subjects in this study comprise LOAD cases and elderly controls aged between 60 and 80 years (at the time of diagnosis, examination or death) that were included in stage one of a LOAD GWAS published in 2009 [3]. The subjects are sampled from three North American Caucasian case control series comprised of clinical subjects from Jacksonville Florida (Mayo-JS series, Neill Graff-Radford MD), Rochester Minnesota (Mayo-RS series, Ron Peterson MD) and an autopsy series collected at the Mayo Clinic Jacksonville (Mayo-AUT series, Dennis Dickson MD). The study was approved by the relevant institutional review board and all participants gave informed consent. The subjects from the stage one GWAS described by Carrasquillo.M.M *et al.* that were included in this candidate gene study are summarized in Table 4.3. Samples were collected and DNA was isolated as described elsewhere [3].

Series	N (AD)	N (control)	Females (%)	Mean Age*	E4 allele (%)
JS	334	307	387 (60%)	73.6	321 (50%)
RS	234	656	466 (52%)	73.9	340 (38%)
AUT	220	212	186 (43%)	72.6	187 (43%)
All	788	1175	1039 (53%)	73.6	888 (45%)

Table 4.3. **Demographic data for the case control series used in the candidate gene association study using logistic regression analysis**. * Age at diagnosis.

Genotyping and Quality control: The subjects in this study were genotyped for over 300,000 SNPs using the HumanHap300-Duo Genotyping BeadChip (Illumina), genotypes were scored using Illumina's bead studio 2.0 software and following removal of SNPs with a call rate < 90% additional quality control measures were executed using PLINK [381] as described by Carrasquillo.M.M *et al.*.

Selection of genes and SNPs: The KEGG website [313-315], which provides pathway data, was used to identify the nuclear genes that are responsible for encoding subunits of the electron transport chain complexes. Genes known to encode subunits of complex IV (cytochrome c oxidase) were selected for this study which led to the selection of 17 genes. Pseudogenes were not included in the study, however genes that encode variant isoforms of a subunit do exist and were included. Genes known to form part of the mitochondrial polymerase complex were also selected: *POLG, POLG2* and *PEO1*. In total, 20 genes were selected for the candidate gene study: the base pair positions based upon Build 36.1 of the human genome were identified using the UCSC genome browser [395] and for each locus

the gene and 20kb 5' and 3' of the gene was designated as the target area for SNP association. Details of each locus and base pair boundaries are detailed in Table 4.4.

Statistical methods: Carrasquillo M.M *et al.* previously generated lgen, map and fam files using PLINK software (provided by Dr Shaun Purcell at http://pngu.mgh.harvard.edu/purcell/plink/), which were used for candidate gene analysis in this study. A covariate file was generated with Sex, Age (at diagnosis/exam: JS and RS series, or death: AUT series) and a dichotomous variable for presence/absence of an *APOE* ε4 allele, to facilitate the inclusion of these covariates in the analysis. PLINK was used to execute logistic regression analysis, using an additive model with covariates; SNPs were specified using the --from-bp and --to-bp PLINK functions to select SNPs for analysis at each of the loci outlined in Table 4.4.

Gene	Chr	Gene Position-Start	Gene Position-End	Gene Size (bp)	Target Locus - Start	Target locus -End
COX4I1	16	84,392,312	84,398,075	5,763	84,372,312	84,418,075
COX4I2	20	29,689,352	29,696,461	7,109	29,669,352	29,716,461
COX5A	15	72,999,670	73,017,548	17,878	72,979,670	73,037,548
COX5B	2	97,628,953	97,631,089	2,136	97,608,953	97,651,089
COX6A1	12	119,360,287	119,362,912	2,625	119,340,287	119,382,912
COX6A2	16	31,346,556	31,347,222	666	31,326,556	31,367,222
COX6B1	19	40,830,995	40,841,524	10,529	40,810,995	40,861,524
COX6B2	19	60,556,777	60,558,275	1,498	60,536,777	60,578,275
COX6C	8	100,959,548	100,975,071	15,523	100,939,548	100,995,071
COX7A1	19	41,333,664	41,335,611	1,947	41,313,664	41,355,611
COX7A2	6	76,004,223	76,010,245	6,022	75,984,223	76,030,245
COX7A2L	2	42,431,148	42,449,654	18,506	42,411,148	42,469,654
COX7B	X	77,041,617	77,047,537	5,920	77,021,617	77,067,537
COX7B2	4	46,431,721	46,431,966	245	46,411,721	46,451,966
COX7C	5	85,949,540	85,952,339	2,799	85,929,540	85,972,339
COX8A	11	63,498,655	63,500,591	1,936	63,478,655	63,520,591
COX8C	14	92,883,290	92,884,453	1,163	92,863,290	92,904,453
POLG	15	87,660,540	87,679,030	18,490	87,640,540	87,699,030
POLG2	17	59,912,173	59,923,631	11,458	59,892,173	59,943,631
PEO1	10	102,737,302	102,744,148	6,846	102,717,302	102,764,148

Table 4.4. Twenty candidate genes selected for this study. gene positions and locus positions are shown based upon build 36.1 of the human genome.

4.3. Results

4.3.1 Mitochondrial mutational burden.

Amplification of nuclear pseudogenes: The PCR reactions were all tested for amplification of nuclear pseudogenes using a DNA sample from human osteosarcoma 143b cell line which contained no mitochondria: ρ^0. Visualization of PCR products for this sample on a 2% TBE ethidium bromide (EtBr) gel was negative in almost all reactions tested, with the exception of the PCR for the *ND1* locus using the Phusion polymerase; under these PCR conditions a faint visible band was occasionally present. However this band had a higher molecular weight than the target PCR product and was not visible in the lanes on the gels with PCR product generated from the experimental samples. It is likely that this non-specific PCR product is not amplified to high levels in the experimental samples as the copy number of mitochondrial target is much higher than any nuclear target. Also, if a nuclear numt had been amplified this would have been detected in the sequencing analysis since any clones containing this product would be longer than the target sequence and would not align to the target sequence correctly. Therefore it is unlikely that clones of this nuclear pseudogene(s) were included in the sequencing analysis.

Low fidelity PCR polymerases artificially inflate the detected mitochondrial mutational burden: For the initial pilot study, testing three PCR polymerases at two mitochondrial loci, the average number of clones with good quality sequence evaluated using the Phusion polymerase was 81, using Platinum Taq polymerase was 76 and using AmpliTaq Gold polymerase was 73. The average number of base

pairs that were analysed per clone for Phusion polymerase was 530bps, Platinum Taq polymerase was 516 bps and Amplitaq Gold was 449 bps. Investigation of the mitochondrial mutational burden at both the *HVR* and *ND1* loci revealed increased mutational burden rates when lower fidelity PCR polymerases were used in the initial PCR step (Table 4.5) which is probably due to a higher number of PCR errors generated when using a lower fidelity enzyme. Analysis of the total mutational burden across the three PCR polymerases using the Kruskal-Wallis test resulted in a p-value of 0.027 at the *HVR* locus and 0.051 at the *ND1* locus, indicating that the measured mutational burden for at least one of the polymerases is significantly different from at least one of the others. Together this data implies that the PCR products generated by the Phusion polymerase were of the best quality for sequence analysis, both in terms of the amount of good quality sequence and low PCR error rate. Consequently, the Phusion polymerase was used in the subsequent study for evaluating the relationship between mitochondrial mutational burden at the *HVR* locus and age in DNA isolated from blood.

The HVR locus has a higher mutational burden than the ND1 locus: Comparison of the total mutational burden at the *HVR* and *ND1* loci of the mitochondrial (mt) genome using Phusion polymerase (Table 4.5) confirmed that the hyper-variable region of the mt genome has a higher frequency of mutation than the *ND1* locus. The total mutational burden at the *HVR* locus is 135 variants/1x10^6 bp, whilst the *ND1* mean mutational burden is 38 variants/1x10^6bp, revealing a 3.6 fold difference in total mutational burden at these loci. The *HVR* is the only non-coding region of the mitochondrial genome and appears to have a higher tolerance for somatic

mutation and hence is called the hyper-variable region. This result is a good positive control for the approach used and indicates that this method is able to detect expected differences in mutational burden.

Mitochondrial mutational burden of the HVR and ND1 loci varies with age. Since it has been demonstrated that the Phusion polymerase has the lowest PCR error rate, the data generated from this polymerase was used for a crude evaluation of the relationship between mitochondrial mutational burden and age in these samples. The results summarized in Table 4.5 show that the mutational burden at both loci is greatest in the 59 yr old sample: 203 variants/1x10^6 bp at the *HVR* locus and 79.6 variants/1x10^6 bp at the *ND1* locus. The lowest mutational burden at the *HVR* locus was observed in the cord blood sample with 44.1 variants/1x10^6 bp, whilst the 97 yr old sample exhibited the lowest mutational burden at the *ND1* locus with 11.7variants /1x10^6 bp. The data imply that mutational burden may vary with aging or from one subject to another; the analysis of additional samples is needed to make more reliable conclusions about any association between aging and mitochondrial mutational burden.

		0 Years			59 Years			97 Years			All		
Locus	**Polymerase**	**#somatic variants**	**bps**	**burden x 10^6**	**#somatic variants**	**bps**	**burden x 10^6**	**#somatic variants**	**bps**	**burden x 10^6**	**#somatic variants**	**bps**	**burden x 10^6**
HVR	Phusion	4	90,618	44.10	15	73,942	203.00	15	87,113	172.00	34	251,673	135.00
	Platinum	30	66,930	448.00	25	74,045	338.00	23	54,120	425.00	78	195,095	400.00
	Gold	29	60,390	480.00	32	47,310	676.00	33	60,701	544.00	94	168,401	558.00
ND1	Phusion	2	89,724	22.30	7	87,989	79.60	1	85,756	11.70	10	263,469	38.00
	Platinum	33	91,437	361.00	26	98,430	264.00	35	86,062	407.00	94	275,929	341.00
	Gold	41	93,889	437.00	28	74,189	377.00	23	57,117	403.00	92	225,195	409.00

Table 4.5. Mitochondrial mutational burden of *ND1* and *HVR* loci for three subjects measured using three DNA polymerases. The somatic mutational burden (somatic variants/base) is defined as the number of somatic variants/total number of bases analysed multiplied by 1E06.

The relationship between age and mitochondrial mutational burden at the HVR locus is unclear: To further evaluate the affect of age on mt mutational burden at the *HVR* locus, 10 additional control samples were selected and grouped with the 59yr old and 97 yr old samples used previously, to generate 4 age groups with 3 samples in each group (Table 4.1). Using the PCR-Cloning method used for the previous analysis, the *HVR* locus was investigated for mt mutational burden in these 12 samples. Table 4.6 shows that the highest mutational burden was observed in the 80-year-old age group: median 88.23 variants/1x10^6 bp and the lowest mutational burden was observed in the 60 year-old age group with a median of 22.41 variants/1x10^6 bp. There is a lot of within group variability of mutational burden where the range of the 3 values is greater than the median value in all groups except for the most elderly group of subjects aged over 90 years. The results do not indicate a direct relationship between age and mutational burden at the *HVR* locus and analysis of variability in mutational burden across the four groups yielded a Kruskal-Wallis p-value of 0.092, indicating a trend for at least one of the groups to yield a larger number of somatic mutations than at least one other group, however this observation does not achieve significance at $p < 0.05$.

The number of clones analysed per sample may not have a linear correlation with the number of somatic mutations identified: Two study samples were included in both the pilot study and the follow-up study (Sample 4, aged 59 years and sample 12, aged 97 years: Table. 4.1). The number of clones selected in the first study was 95, of which 48 were randomly selected for inclusion in the second study. The number of somatic variants identified in the first study was 15 for each of the

samples. However, in the second study, only 1 and 2 somatic variants were identified for the 59 and 97 year old samples respectively. If there was a linear relationship between the number of clones analysed and the number of variants identified then it would be expected that between 7 and 8 somatic variants would have been identified in the analysis of 48 clones. It is therefore possible that this study is underpowered to detect differences between age groups based upon picking only 48 colonies per sample: many more colonies may need to be selected for the study.

Group	Sample #	Age	No. Somatic Variants	Bps	burden $x10^6$	Median burden $x10^6$	Range
A	1	43	2	29,293	68.28	68.28	72.27
	2	46	4	39,207	102.02		
	3	46	1	33,614	29.75		
B	4	59	1	44,619	22.41	22.41	49.0
	5	60	0	39,369	0.00		
	6	60	2	40,817	49.00		
C	7	80	7	43,714	160.13	88.23	109.98
	8	80	4	45,335	88.23		
	9	80	2	39,878	50.15		
D	10	92	1	35,326	28.31	47.46	20.38
	11	93	2	41,076	48.69		
	12	97	2	42,143	47.46		

Table 4.6. Mitochondrial mutational burden for 12 control samples Samples have a range of ages, organized into 4 age groups.

Overall, these data indicate that the mutational burden at the HVR locus in leukocytes is highly variable and may not be influenced by age. Further analysis with a much larger number of samples, colonies and age groups is likely to be

necessary to establish reliable estimates of the mean mutational burden for each age category. This type of study would likely be cost prohibitive using the approach described here due to the large numbers of colonies that would need to be sequenced, so alternative sequencing methods would need to be explored. For these reasons, the measurement of mitochondrial mutational burden was not explored further in DNA isolated from the target tissue: the human brain.

An alternative to measuring mutational burden is to measure DNA damage: various published approaches were considered and a method using absolute quantification by qPCR was selected for evaluation in mtDNA isolated from brain tissue, the results of this study are described over the next pages.

4.3.2 Mitochondrial DNA damage.

The nDNA (RNaseP and β-actin) measurements are highly correlated: Two nuclear loci were measured in each sample to improve the reliability of nuclear quantification. The measured values were highly correlated: cerebellum: $r = 0.986$, $p < 0.001$, temporal cortex: $r = 0.991$, $p < 0.001$, indicating that the two sets of values are very similar. The mean of the nuclear quantification estimates for each of the two nuclear loci was used to determine the proportion of nDNA in each sample.

The mtDNA measures are highly correlated: Three mitochondrial loci were measured using qPCR and pair wise comparisons were made to determine if there was any correlation between the measured values in the two sets of tissues studied. Remarkably, there was a high level of correlation between each pair of measures taken for each set of 10 samples based upon diagnosis and tissue (cerebellum and

temporal cortex), the correlation coefficients are detailed in Table 4.7. This result indicates that, in these subjects, there is not a large difference in the relative levels of measurable mtDNA at different areas of the mt genome.

Tissue/Dx	Assay: mean mtDNA damage (rank)^				Correlation coefficient*		
	16s	D-loop	COII	P-value	16s-D-loop	16s-COII	D-loop-COII
Cer -AD	90.20 (2.70)	74.80 (1.50)	78.94 (1.90)	0.017	0.96	0.97	0.96
Cer - Normal	87.13 (3.00)	67.82 (1.70)	70.25 (1.30)	<0.001	0.90	0.97	0.89
Cer -PSP	86.91 (3.00)	57.90 (1.70)	62.66 (1.30)	<0.001	0.91	0.97	0.93
Cer -VaD	97.78 (2.30)	94.65 (1.80)	95.49 (2.00)	0.461	0.99	0.99	1.00
Tx -AD	96.09 (2.70)	90.89 (1.50)	92.30 (1.80)	0.020	0.99	0.95	0.92
Tx - Normal	95.21 (2.90)	88.55 (1.20)	91.14 (1.90)	<0.001	0.97	0.96	0.97
Tx -PSP	91.72 (3.00)	77.54 (1.30)	79.99 (1.70)	<0.001	0.98	0.99	0.97
Tx -VaD	97.53 (2.60)	95.16 (1.70)	95.81 (1.70)	0.067	0.95	0.98	0.99

Table 4.7 Absolute levels of mtDNA damage measured for 4 pathological categories and regions of the brain for three mitochondrial loci.^ Friedmans test, the average rank for each assay is presented in parenthesis. *Correlation in Excel

An evaluation of the DNA damage data indicates that, overall, the *16s* locus consistently exhibited the greatest levels of damage across all disease groups. The average mtDNA damage at the *16s* locus is estimated at 92.82%, whereas the corresponding values for *COII* and *D-loop* assays are 83.32% and 80.92% respectively. The results of Friedman's test indicate that the absolute levels of damage at one or more of the loci is significantly different from at least one other

locus in all except the VaD subjects, where the p-value was >0.05 in both cerebellum and temporal cortex (Table 4.7). This is probably due to the fact that the levels of damage in the VaD tissue are very high at all three mitochondrial loci and so the difference between the damage values at the different loci is much smaller. So despite the levels of mtDNA at each locus being significantly correlated in pair wise comparisons, the assessment of mtDNA damage at each of these loci indicates a significant difference overall between the different mitochondrial loci.

There are significant, tissue-specific, differences between the extent of mtDNA damage in diseased versus normal tissue: To determine the difference between the extent of mtDNA damage in each of the disease groups against the damage occurring as a part of "normal" aging, the ratio of the mean mtDNA damage values between each disease group and the normal group was calculated for each locus and tissue type. The results are summarized in Table 4.8. The significance of the observed differences was determined using the Wilcoxon rank sum test (a.k.a. Mann-Whitney test) where a p-value of less than 0.05 led to the rejection of the null hypothesis that there is no difference between the mtDNA damage values in diseased versus normal tissue.

VaD: The data presented in Table 4.8 indicates that VaD subjects have significantly higher levels of mtDNA damage than normal subjects; this difference is more pronounced in the cerebellum (ratio 1.12-1.40) than in the temporal cortex (ratio 1.02-1.05) at all three loci examined.

PSP: The data in Table 4.8 indicates that subjects with PSP have lower levels of mtDNA damage than that found in normal aging. This difference is significant in the temporal cortex but not in the cerebellum indicating some tissue selectivity of mtDNA damage in PSP subjects.

AD: The data in Table 4.8 shows that whilst the ratio of mtDNA damage in AD vs normals indicates a trend for an increase in damage in AD's, this difference is not significant in either the cerebellum or temporal cortex tissues (p-value 0.211 to 0.968). This result is consistent across all three mitochondrial loci and implies that mitochondrial DNA damage, at the loci examined, may not play a significant role in LOAD disease aetiology. This is in contrasts to the data for the PSP and VaD subjects and implies that mitochondrial DNA damage is complicit in other neurodegenerative pathologies.

The temporal cortex exhibits greater levels of mtDNA damage than the cerebellum in normal aging and disease: To determine the difference between the levels of mtDNA damage measured in the temporal cortex vs the cerebellum, the ratio of mean mtDNA damage values between the two tissues was determined for each phenotype and mtDNA locus. The Wilcoxan signed-rank test was used to determine if differences observed were significant, where a p-value of < 0.05 led to the rejection of the null hypothesis that there is no difference between the mDNA damage values in the temporal cortex when compared with the cerebellum.

Tissue / Assay	Dx	mtDNA damage (%)*				Vs Normal$		Vs Cer#	
		Min	Max	Mean	CV	Ratio	p-value	Ratio	p-value
Cer 16s	Normal	80.13	93.68	87.13	0.05	-	-	-	-
	AD	81.09	99.90	90.20	0.08	1.04	0.430	-	-
	PSP	76.30	91.35	86.91	0.05	1.00	0.734	-	-
	VaD	91.06	99.95	97.78	0.03	1.12	3E-04	-	-
Tx 16s	Normal	91.34	99.73	95.21	0.03	-	-	1.09	0.002
	AD	93.40	99.76	96.09	0.03	1.01	0.675	1.07	0.084
	PSP	86.50	96.38	91.72	0.04	0.96	0.038	1.06	0.010
	VaD	92.38	99.77	97.53	0.02	1.02	0.038	1.00	0.695
Cer COII	Normal	53.10	87.42	70.25	0.16	-	-	-	-
	AD	56.69	99.95	78.95	0.20	1.12	0.211	-	-
	PSP	26.78	81.73	62.66	0.26	0.89	0.447	-	-
	VaD	78.84	99.97	95.49	0.07	1.36	4E-04	-	-
Tx COII	Normal	85.51	99.69	91.14	0.05	-	-	1.30	0.002
	AD	84.92	99.80	92.30	0.06	1.01	0.968	1.17	0.037
	PSP	63.78	94.66	79.99	0.13	0.88	0.014	1.28	0.006
	VaD	85.35	99.84	95.81	0.04	1.05	0.017	1.00	0.922
Cer D-loop	Normal	54.58	82.99	67.82	0.14	-	-	-	-
	AD	47.88	99.95	74.86	0.26	1.10	0.569	-	-
	PSP	31.06	77.57	57.90	0.28	0.85	0.211	-	-
	VaD	74.66	99.93	94.65	0.09	1.40	3E-04	-	-
Tx D-loop	Normal	75.72	99.76	88.55	0.08	-	-	1.31	0.002
	AD	83.06	99.84	90.89	0.07	1.03	0.624	1.21	0.084
	PSP	57.42	93.02	77.54	0.16	0.88	0.038	1.34	0.006
	VaD	78.95	99.82	95.16	0.06	1.07	0.017	1.01	1.000

Table 4.8. Evaluation of mtDNA damage measured at three loci for samples obtained from two regions of the brain and 4 pathological groups. * damage determined as described in section 4.2.22. $ Ratio of damage between the target group and the normal group within the same tissue for each assay, p-value determined using the Wilcoxon rank sum test (a.k.a. Mann-Whitney test), # Ratio of damage between the temporal cortex and the cerebellum for each sat of 10 paired values, p-value calculated using Wilcoxon signed-ranks test.

VaD: The analysis did not identify significant differences between the levels of mtDNA damage in the temporal cortex and cerebellum at any of the three mitochondrial loci. The ratio of damage in temporal cortex versus cerebllum tissue was 1.00 – 1.01 for VaD subjects. This indicates that the difference in the magnitude of the ratio when VaD were compared to normal subjects in the previous analysis is likely due to differences between the cerebellum and temporal cortex in the normals and not in the VaD subjects.

Normal: The results presented in Table 4.8 indicate that at all three mitochondrial loci there is selective mtDNA damage in the temporal cortex, when compared with the cerebellum, during normal aging. The ratio of DNA damage was smallest at the *16s* locus (1.09) but was highly significant across all three loci (p-value = 0.002 at all loci).

PSP: The ratios presented in Table 4.8 indicate that the levels of mtDNA damage are significantly greater in the temporal cortex than in the cerebellum at all three mitochondrial loci for PSP subjects (p-values 0.010 to 0.006). Similar to the normal subjects, this difference was least at the *16s* locus, but still remained significant (ratio 1.06, p-value = 0.010)

AD: The AD subjects exhibited a positive ratio (>1.00) of mtDNA damage between the temporal cortex and the cerebellum at all three mitochondrial loci, that was significant at the *COII* locus (p-value=0.037) but did not achieve statistical significance at the *16s* or *D-loop* loci (p-value=0.084). This finding was unique to the AD subjects, as the other three disease groups investigated exhibited either significant or no significant difference, between the temporal cortex and cerebellum

at all three mitochondrial loci. This indicates that for AD subjects, the *COII* locus might be more sensitive to DNA damage in affected tissues than other mitochondrial loci, when compared to a baseline level of damage in unaffected LOAD tissue. This finding is of relevance to the study of LOAD, given the associations reported between cytochrome c oxidase activity and LOAD.

Sensitive quantification of numts using Rho^0 DNA: The amplification of nuclear mtDNA pseudogenes, also referred to as numts, may be a potential confounder to any study of mitochondrial DNA that involves PCR amplification. Throughout this chapter the use of ρ^0 DNA, isolated from cells with no mitochondria, has been described as a way to assess the tendency of any PCR reaction to amplify numts. Previously, this has been assessed by running PCR products on an ethidium bromide gel and visualizing the presence or absence of visible bands from the ρ^0 sample. However, the use of qPCR is a more sensitive method for identifying amplification of numts. For each of the three qPCR mitochondrial assays, a ρ^0 DNA sample was also measured, in triplicate, on the same plate as the experimental samples, to determine if the assays were measuring mtDNA specifically or if there was a contribution from nDNA pseudogenes or numts. The ρ^0 DNA was plated at a spectrophotometrically measured concentration of 2ng/ul on the plates used to assay the mitochondrial loci. For all three of the mitochondrial assays, the ρ^0 sample had increased levels of product amplification when compared with the no template control; however this varied depending on the assay.

The ratio between the number of copies of amplified "genes" measured in the ρ^0 sample when compared with the no template control (background) is a good

measure of how specific the qPCR assay is to mitochondrial DNA sequence. For the *D-loop* assay this ratio was small, at 1.08, however the *16s* and *CoxII* assays had much higher ratios of 117 and 10 respectively. Given this data, it is highly unlikely that the *D-loop* assay targets any nuclear locus; however both the *16s* and *CoxII* assays do show evidence of amplification of numts.

For the *CoxII* assay, the measured amount of [mtDNA] for the ρ^0 sample was greater than 1 of the experimental samples, but was 255 times smaller than the average measured value for all of the experimental samples. For the *16s* assay, the measured amount of [mtDNA] for the ρ^0 sample was greater than 2 of the experimental samples. The data at the *16s* and *CoII* loci may therefore be confounded by amplification of numts. Consequently the results from the *D-loop* may be the only mitochondrial assay that can be reliably interpreted. However, it is important to consider the fact that the ρ^0 sample is 100% nuclear DNA and so the data from this sample is an extreme test of the tendency for PCR primers to amplify nuclear DNA. In an experimental sample the majority of the DNA is mitochondrial DNA and there are many more copies of the mitochondrial genome per ng of DNA than copies of the nuclear genome per ng of DNA, resulting in the number of mitochondrial targets far outweighing the number of nuclear targets. This likely reduces the impact of amplification of numts on the data obtained from the experimental samples. Furthermore the *CoII* and *16s* data are highly correlated with the *D-loop* data; this would be unlikely to occur if the *CoII* and *16s* assays were significantly influenced by the amplification of numts. Consequently, whilst the results obtained from the

CoII and *16s* assays need to be interpreted with caution; it is likely that the reduced specificity of these assays may not have a significant impact on the data.

For the analyses described on the previous pages, the p-values have not been corrected for multiple tests; however given the small sample size and the small number of tests for each hypothesis tested, a p-value of <0.05 was used to reject the null hypothesis. This may lead to a small number of false-positive findings, however the goal of this study is to function as a pilot study and any conclusions drawn from the study would need to be confirmed in additional studies with a larger number of samples.

4.3.3 Candidate gene study.

Previously in the chapter, the effect of damage to the mitochondrial genome was specifically investigated using techniques to measure both mutational burden and DNA damage. The mitochondria however, rely on both nuclear and mitochondrial genes to function and so the study of nuclear genes that encode key proteins necessary for mitochondrial function is also of relevance in the study of mitochondrial genetics in dementia aetiology. To this end, the mitochondrial polymerase (polymerase γ) which functions in mitochondrial DNA replication and integrity is an excellent candidate gene, as it is encoded by nuclear genes and is the only know DNA polymerase to function in the mitochondria. Three genes: *POLG*, *POLG2* and *PEO1* form some of the necessary components for mitochondrial DNA replication and so were investigated for association with LOAD using data from a published LOAD GWAS [3].

The results of this study (Table 4.9) identify two SNPs which show nominally significant association with LOAD: rs2351000 at the *POLG l*ocus which shows significantly increased risk for LOAD associated with the minor allele (OR: 1.36, p-value=0.006) and rs1991401 at the *POLG2* locus which shows significantly decreased risk for LOAD associated with the minor allele (OR=0.84, p-value=0.025). Neither of these SNPs maintain significance after a Bonferroni correction for multiple tests, however an association study of 18 SNPs would be expected to yield less than 1 SNP with a p-value < 0.05 by chance and so both of these SNPs warrant further study in additional LOAD case control series. This is considered stage 1 of a candidate gene study where follow-up in additional samples would be needed to confirm any identified associations.

Gene	Chr	SNP	Position	N	OR	L95	U95	P-value
POLG	15	rs1138465	87,659,606	1940	0.91	0.79	1.05	0.181
		rs976072	87,660,381	1959	0.91	0.79	1.05	0.203
		rs3087374	87,660,998	1960	1.07	0.84	1.36	0.576
		rs2351000	**87,670,737**	**1922**	**1.36**	**1.09**	**1.69**	**0.006**
		rs2247233	87,676,941	1957	1.01	0.88	1.16	0.905
		rs7495044	87,685,173	1956	1.05	0.91	1.21	0.503
		rs176641	87,691,186	1947	1.12	0.96	1.29	0.146
POLG2	17	rs1427463	59,923,044	1962	1.15	0.91	1.45	0.253
		rs1140409	59,927,132	1956	0.89	0.67	1.17	0.395
		rs2075552	59,931,215	1947	1.21	0.92	1.58	0.179
		rs1991401	**59,932,897**	**1952**	**0.84**	**0.73**	**0.98**	**0.025**
PEO1	10	rs2863095	102,736,493	1935	0.92	0.78	1.09	0.341
		rs3740484	102,737,353	1931	0.96	0.83	1.11	0.575
		rs3740487	102,740,773	1938	0.91	0.77	1.08	0.301
		rs701836	102,757,107	1894	0.97	0.83	1.15	0.742
		rs807023	102,760,072	1907	1.15	0.94	1.40	0.175
		rs11190790	102,762,273	1949	1.01	0.87	1.16	0.939
		rs807019	102,762,830	1948	1.06	0.88	1.28	0.558

Table 4.9. Single SNP association at three loci. rs2351000 is located in intron nine of the *PolG* gene, 110bp 3' of exon nine. **rs1991401** is located in the 5'UTR of the gene *DDX5* and is 9 kb 3' of the *POLG2* gene.

Decreased activity of cytochrome c oxidase (complex IV of the mitochondrial electron transport chain) has been frequently associated with Alzheimer's disease. The catalytic subunits of this multi-subunit enzyme are encoded by three genes in the mitochondrial genome (COXI-III), with the remainder of the subunits being encoded by the nuclear genome. Whilst the nuclear encoded subunits do not have catalytic functions, they undoubtedly play an important role in the function of this enzyme. Consequently, the investigation of SNPs at the nuclear genes that encode COX subunits is of relevance to the study of this enzyme in LOAD aetiology. The human cyctochrome c oxidase enzyme is composed of 13 subunits, 10 of which are

nuclear encoded; however tissue specific isoforms exist for several of the subunits. In this study, genes for all 10 subunits and 7 non-pseudogene isoforms were included in the analysis. SNPs in these genes were evaluated using data available from the Carrasquillo *et al*. LOAD GWAS [3].

The result of this study identified four SNPs with nominally significant association with LOAD in two of the seventeen genes investigated. None of the SNPs would maintain significance after a Bonferroni correction for multiple tests, however like the previous study there are a greater number of SNPs with a p-value of less than 0.05 than would be expected by chance (3.5 by chance, 4 observed), consequently all four SNPs warrant further investigation in additional LOAD case control series.

One of the genes, *COX7B* did not have any SNPs within the defined locus and is therefore not included in the results tables.

Two genes: COX6B1 and COX6B2, which encode two isoforms of the COX6B subunit, harbor SNPs at their respective loci that show nominally significant associations with LOAD, whereas none of the other loci harbor SNPs with significant association (Table 4.10). Both of the COX6B genes are found to be expressed in brain and both are also located on chromosome 19, approximately 10 million base pairs from the *APOE* locus.

Gene	Chr	SNP	Position	N	OR	L95	U95	P-value
COX4I1	16	rs6120890	29,671,320	1958	1.03	0.86	1.24	0.752
		rs6060403	29,684,765	1939	1.07	0.91	1.25	0.418
		rs11907253	29,696,334	1959	1.09	0.85	1.4	0.493
		rs6060454	29,696,565	1954	1.11	0.87	1.42	0.415
		rs6058259	29,698,963	1947	1	0.86	1.16	0.981
		rs6060531	29,707,019	1952	1.11	0.86	1.42	0.436
		rs6060563	29,712,464	1961	0.9	0.77	1.05	0.164
COX4I2	20	rs11860152	84,374,825	1961	0.78	0.57	1.06	0.115
		rs896253	84,377,703	1961	0.76	0.56	1.05	0.096
		rs10514609	84,381,574	1957	0.91	0.77	1.07	0.245
		rs2291658	84,384,374	1908	1.02	0.85	1.24	0.81
		rs2733954	84,393,757	1959	0.92	0.77	1.1	0.376
		rs11117407	84,411,812	1959	0.95	0.8	1.14	0.591
COX5A	15	rs4886636	72,983,229	1913	1.06	0.92	1.22	0.408
		rs11856413	72,986,945	1956	1.07	0.93	1.23	0.341
		rs11072518	73,021,663	1954	1.05	0.91	1.22	0.489
COX5B	2	rs1800649	97,631,289	1957	1.05	0.91	1.21	0.512
		rs11692435	97,641,786	1946	1.09	0.85	1.39	0.498
COX6A1	12	rs206965	119,340,715	1958	1.06	0.89	1.26	0.505
		rs4766961	119,346,915	1950	0.98	0.84	1.13	0.74
		rs2235218	119,373,002	1945	0.97	0.84	1.13	0.721
		rs7137953	119,379,068	1958	1.15	0.96	1.37	0.14
COX6A2	16	rs8056122	31,335,179	1957	0.96	0.83	1.1	0.536
		rs4889657	31,344,433	1955	0.84	0.62	1.13	0.238
		rs8045738	31,357,499	1957	1.03	0.89	1.19	0.703
COX6B1	19	rs2285415	40,814,281	1943	0.97	0.85	1.12	0.697
		rs2285419	**40,826,048**	**1948**	**0.78**	**0.61**	**1**	**0.046**
		rs3761093	40,828,289	1928	1.14	0.99	1.32	0.078
		rs7254601	40,839,155	1953	1.12	0.95	1.31	0.176
		rs4806187	40,841,601	1948	0.99	0.86	1.14	0.894
		rs2267584	40,848,428	1944	0.98	0.84	1.14	0.786
		rs2267586	**40,851,208**	**1958**	**0.72**	**0.57**	**0.91**	**0.006**
COX6B2	19	rs7253685	60,537,198	1952	0.9	0.78	1.04	0.141
		rs1870073	60,542,081	1957	0.97	0.83	1.13	0.662
		rs2384685	60,546,125	1880	0.93	0.81	1.07	0.306
		rs1058511	60,551,251	1930	1.04	0.91	1.2	0.545
		rs11084396	60,557,339	1949	1.07	0.93	1.22	0.346
		rs12976922	60,562,163	1954	1.06	0.87	1.3	0.567
		rs897799	60,564,003	1937	1.15	0.97	1.37	0.119
		rs1126757	**60,571,684**	**1947**	**0.83**	**0.73**	**0.96**	**0.01**
		rs6509940	**60,575,691**	**1938**	**1.18**	**1.03**	**1.36**	**0.02**

Gene	Chr	SNP	Position	N	OR	L95	U95	P-value
COX6C	8	rs4626565	100,959,718	1953	1.14	0.95	1.37	0.154
		rs4518636	100,973,453	1961	1.13	0.94	1.36	0.182
		rs13277184	100,986,128	1945	1.12	0.92	1.36	0.249
		rs10098852	100,990,479	1948	0.98	0.84	1.13	0.735
		rs7837596	100,993,447	1958	1.13	0.91	1.4	0.264
COX7A1	19	rs10410162	41,319,630	1958	1.11	0.96	1.28	0.162
		rs7255180	41,332,159	1948	1	0.74	1.35	0.992
		rs753420	41,335,535	1931	0.87	0.75	1.02	0.09
		rs11665903	41,344,025	1939	1	0.86	1.17	1
COX7A2L	2	rs6716767	42,412,760	1960	0.91	0.73	1.13	0.38
		rs1981664	42,430,674	1942	0.97	0.84	1.12	0.675
		rs10206058	42,437,171	1952	1.03	0.89	1.19	0.693
		rs1992258	42,453,421	1870	1.08	0.94	1.25	0.28
		rs7592036	42,457,613	1949	1.1	0.95	1.28	0.189
		rs12615383	42,464,055	1961	0.99	0.81	1.2	0.909
		rs7560797	42,469,000	1959	0.98	0.84	1.14	0.751
COX7A2	6	rs431116	76,003,142	1961	1.16	0.89	1.52	0.27
		rs654428	76,020,471	1953	1.12	0.81	1.54	0.506
COX7B2	4	rs9790574	46,429,658	1962	0.98	0.82	1.17	0.799
		rs4395475	46,437,406	1944	1	0.81	1.24	0.999
COX7C	5	rs4605828	85,942,414	1938	1.11	0.97	1.27	0.136
		rs10805837	85,947,410	1959	1.05	0.91	1.21	0.497
		rs17476583	85,958,923	1957	0.98	0.84	1.14	0.803
COX8A	11	rs11231657	63,482,185	1953	0.99	0.7	1.4	0.942
		rs481835	63,482,358	1945	0.95	0.83	1.1	0.513
		rs480211	63,496,015	1955	0.98	0.86	1.13	0.823
		rs483340	63,511,562	1936	1	0.87	1.15	0.968
COX8C	14	rs7155081	92,869,929	1957	1.15	0.91	1.46	0.229
		rs6575325	92,898,942	1962	1.04	0.89	1.21	0.631

Table 4.10. Single SNP association for 16 genes that encode components of the cytochrome oxidase complex.

***COX6B1*:** Two SNPs at the *COX6B1* locus showed nominally significant association with LOAD in the case control series tested. rs2285419 (OR 0.78, p-value 0.046) is a coding SNP (Asp to Asn substitution) in exon 5 of the gene *ETV2* and is 19kb 5' of the *COX6B1* gene. The second SNP, rs2267586 (OR 0.72, p-value 0.006) is also a coding SNP (Ser to Thr substitution) in exon 2 of the gene *UPK1A* and is 10kb 3' of the *COX6B1* gene.

***COX6B2*:** Two SNPs at the *COX6B2* locus showed nominally significant association with LOAD in the case control series tested. rs1126757 (OR 0.83, p-value 0.010) is a synonymous SNP (Ala to Ala) in exon 3 of the gene *IL11* and is 14kb 3' of the *COX6B2* gene. The second SNP, rs6509940 is 18kb 3' of the *COX6B*2 gene.

It is worth noting that the SNP's identified at the two *COX6B* loci do not lie within either gene but are greater than 10kb away and in some cases within the exons of flanking genes. It is therefore unlikely that the SNPs identified in this study have a functional influence on the *COX6B* genes, but are rather in LD with functional variants. It is also possible that these findings are influenced by *APOE* or are in LD with each other, however given the physical distance in bps between the two loci (~2 million bps) and with *APOE (*~1 million bps), this is unlikely.

The data presented here implies a possible role for the COX6B subunit of complex IV of the ETC in LOAD aetiology, further investigation of these loci with additional SNPs and case control series is needed to both confirm the detected association and to fine map the loci.

4.4. Summary

The role of mitochondrial genetics in the aetiology of LOAD has yet to be elucidated. In this chapter, various approaches for evaluating the role of mitochondrial genetics are tested and preliminary data highlights the need for additional studies in this area of research.

It is beyond the scope of this chapter to explore all possible hypotheses for the role of mitochondrial genetics in LOAD and so two aspects of mitochondrial genetics were focused on: i. Methods for evaluating mitochondrial mutational burden and DNA damage and ii. A candidate gene study of nuclear genes selected for their importance in mitochondrial function.

Initially a pilot test was designed to study mitochondrial mutational burden using a previously reported method (based upon PCR, cloning and sequencing), using DNA isolated from blood leukocytes in order to evaluate: (A) the effectiveness of the method for detecting differences in mitochondrial mutational burden and (B) the validity of using DNA isolated from circulating cells, which are readily available and can be collected from live subjects, to detect changes in mitochondrial burden with age. If age related changes can be detected in DNA isolated from blood then it may be possible to explore the use of mitochondrial mutational burden as a clinical diagnostic tool for evaluating risk for age related diseases.

The pilot study confirmed the need for a high fidelity polymerase when employing PCR based methods for studying mitochondrial mutational burden. The PCR error rate of the highest fidelity polymerase used in this study, Phusion, is reported to be 4.4E-07 errors/bp (or 0.44 errors per million base pairs), much lower than the

detected mutational burden for the samples investigated using this polymerase (average 57 mutations per million base pairs). This indicates that, after correction for PCR errors, the loci investigated in this study have a detectable level of mutational burden. The analysis of mitochondrial mutational burden in additional samples, representing an expanded range of ages was unable to establish a clear relationship between age and mitochondrial mutational burden. Whilst the approach tested here for measuring mitochondrial mutational burden was validated in the tissue tested (white blood cells), the approach appears to require a large number of clones to be sequenced and is therefore financially prohibitive for the study of many samples.

In addition to evaluating methods for detecting changes to mitochondrial DNA in blood, this chapter also explored mitochondrial DNA changes in the brain. Given that the previous mutational burden approach identified the need for a large number of samples and clones to be sequenced, an alternative method was employed for exploring mtDNA isolated from brain tissue. Here the use of qPCR was tested as a means of measuring mtDNA damage by adapting an approach used previously by others [386]. The results of the DNA damage study indicate that mtDNA damage may be a major player in VaD aetiology, but the data is far less convincing for Alzheimer's disease. In AD samples, the most interesting result was a significant increase in mtDNA damage at the *COII* locus in the temporal cortex compared with the cerebellum. This comparison was not significant for the other two loci tested, although the trend was in the same direction. This might indicate that *COII* is more damaged in the temporal cortex in particular compared with other mitochondrial loci.

The final study described in this chapter was a candidate gene study for association between Alzheimer disease and genetic variants in nuclear genes that encode proteins essential for optimal functioning of the mitochondria. Genes were selected that are involved in mitochondrial DNA replication since defects in this process may lead to increased incorporation of somatic mutations and decreased mtDNA integrity. Additionally, nuclear genes that encode subunits of the terminal enzyme complex of the electron transport chain, complex IV, were also investigated, since deficiencies in this enzyme have previously been linked to LOAD.

Nominally significant association was identified for six SNPs included in the study, which is more than would be expected by chance alone; indicating that at least one of these SNPs is a true positive association. However the p-values for these six SNPs would not survive a bonferonni correction for multiple tests. This data cautiously indicates a possible role for genes involved in mtDNA replication and in particular subunit COX6B of the cytochrome c oxidase, with LOAD. Genotyping of the SNPs identified in the study reported here warrant further analysis in additional samples to better characterise these associations.

Overall this chapter outlines a novel approach for evaluating the mtDNA damage and provides insight into the extent of mtDNA damage that can be measured from frozen brain tissue samples of different neurodegenerative diseases. This study also identifies two novel candidate LOAD loci proximal to the genes COX6B1 and COX6B2 that could possibly impact risk for LOAD through alterations in cytochrome c oxidase function.

Chapter 5. Discussion

The majority of Alzheimer's disease occurs in subjects older than age 65. The importance of research into the molecular mechanisms and causes underlying AD is of increasing significance given the reality of an aging society.

This thesis presents novel approaches for studying the genetic and biological risk factors that can influence AD and provides a useful resource for future studies. The key findings for each of the studies presented in this thesis were summarised at the end of the results chapters (2, 3 and 4). These findings are briefly reviewed over the next pages and discussed in the context of their impact on AD research.

5.1 AD facts and figures.

As a means to following and documenting the impact of AD on society, Alzheimer's disease international (ADI) publishes an annual report summarising the global costs of this disease and the number of people it affects. The most recent numbers indicate that AD costs $604 billion globally and there are over 30 million people in the world currently suffering from AD, with approximately 4.6 million new cases every year [396]. These numbers are expected to increase dramatically over the next 50 years due to improvements in health care and aging populations [397, 398].

Currently the only medications available to AD sufferers treat only the symptoms and have very little impact on delaying the disease course significantly. There is no cure for AD and there are also no tests (beyond genetic tests for EOFAD) that can predict who will develop the disease prior to the emergence of symptoms. Strategies

are needed for both identifying who is at risk of developing this disease and for producing advances in preventative or curative therapies.

5.2 Overview of AD therapeutics

Current drug therapies: Currently there are five drugs approved in the U.S. for treating the symptoms associated with AD. All of these therapies have been shown in clinical studies to improve the cognitive function of AD patients receiving the drug compared with placebo. The molecular pathways targeted by these drug therapies are: (i) inhibition of acetylcholine breakdown via the administration of acetylcholinesterase inhibitors (Donepezil, ENA-713, Galantamine and Tacrine) or (ii) attenuation of NMDA receptor function (Memantine). Acetylcholine (ACh) is a neurotransmitter that is directly involved in learning and memory functions in the brain. Inhibition of the metabolism of ACh as a treatment for memory dysfunction was explored largely based on the cholinergic hypothesis for neurodegenerative diseases such as AD [399]. In the case of memantine the evidence suggests that therapeutic activity is largely due to the protection of vulnerable cholinergic neurons that may be susceptible to glutamate induced excitotoxicty [400, 401].

Whilst these treatments are fairly well tolerated and provide some relief to AD patients and carers, none of these drugs delay the progression of the disease and the number of patients that respond positively is relatively low [402-405]. Consequently, alternative therapies are critical for preventing an AD epidemic.

Drug trials and future prospects: An improved understanding of the molecular mechanisms that contribute towards the development of AD is critical for the development of therapeutic strategies.

With the identification of Aβ as a major pathological peptide in AD and the subsequent advent of the 'amyloid hypothesis', a large effort was made towards targeting Aβ production and removal for therapeutic purposes. Among the targets explored were inhibition of γ and β-secretase, responsible for the generation of Aβ through APP cleavage, and Aβ immunization therapy was also investigated [406-408]. Unfortunately, many of the drug targets explored have experienced specificity and toxicity problems. More recent approaches aim to avoid these problems, for example multiple Aβ immunization therapies are in clinical trials using a shorter epitope of the Aβ peptide [409, 410]. An updated account of current drug trials is maintained on the Alzforum website (www.alzforum.org).

In this thesis, several approaches for identifying novel drug targets through improved understanding of the genetics of AD were explored. The results of these studies are discussed below.

5.3 Genome-wide screen for novel genes that influence plasma Aβ.

The aim of this project was to identify loci that influence plasma concentrations of Aβ, a key peptide in AD pathogenesis.. The rationale behind this was that genes that influence plasma Aβ are also likely to influence risk for LOAD. The loci identified

in this study represent excellent candidate genes for study in LOAD that may have otherwise gone undetected using a LOAD case-control association study approach. This rationale has previously been validated in extended LOAD pedigrees with extreme plasma Aβ concentrations [62]. In the present study, the populations investigated did not have extreme plasma Aβ levels and were not selected based upon probands with LOAD, but were instead sampled from two isolated populations.

Isolated populations provide several advantages for identifying genetic loci that contribute to phenotypes of interest. For example, extended linkage disequilibrium in these populations can decrease the number of marker genotypes required to adequately "tag" the common haplotypes and can increase the power of the study for detecting association. Furthermore, genetic and environmental heterogeneity is far better controlled in studies of population isolates. However, the study of isolated populations can introduce added complexity requiring the use of statistical approaches that take into account the higher degree of relatedness in the population [411-413]. This was addressed in this study by using the statistical tools available in the R package GenABEL. Most significantly, the tools in this package (mmscore) allow the user to account for the relatedness of study subjects based upon genetic data (IBD) rather than relying on reported relationships. Reported genealogical data can be a potential confounder in these types of studies where inaccuracies in the reporting can occur. Also, when studying a population isolate there may be a higher degree of unknown relatedness (2^{nd} cousins etc) than in outbred populations.

The tendency of a set of samples in an analysis to have inflated test statistics due to confounders such as population substructure is represented by a scaling factor termed lambda (λ). The populations studied in Chapter 2 consist of subjects many of whom are distantly related and a smaller proportion of who are either closely related or unrelated. Consequently the mmscore test (FASTA) was used to execute the GWAS analysis, since it can incorporate the estimated genetic relationships between all pairs of subjects in the study (gkin). The model estimate of λ was generated for each analysis and did not indicate any inflation in the test statistic. On the contrary, for some of the Aβ traits, the value of λ indicated deflation of the test statistic. This effect can similarly be viewed by looking at quantile-quantile (Q-Q) plots that represent the distribution of p-values obtained in the analysis against the null hypothesis of no association beyond what would be expected by chance (Chapter2).

Linkage analysis: Initially, plasma Aβ levels were assessed in the population of Vis using a linkage analysis approach. In order to implement this analysis, a subset of the population was divided into pedigrees of varying size and complexity and IBD estimates were generated using MCMC methods implemented in the program Loki. Linkage analysis was then implemented in Linux using the software package SOLAR [321] which makes use of a variance component approach for quantitative trait analysis. This approach may result in reduced power to detect loci that influence the trait variance since many sampled individuals were not members of known pedigrees, whereas all individuals are used with the genomic relationships

approach implemented in GenABEL, where pedigrees are not defined [414] and the many unknown relationships in the population are accounted for.

The analysis implemented in SOLAR did identify plasma Aβ40 as a significantly heritable trait in the Vis population with an estimated heritability of 16.5%. This is substantially lower than the estimated heritability observed for this trait by Ertekin-Taner *et al.* in extended LOAD pedigrees (~41%) [62, 204]. It is not surprising however that the heritability of this trait is lower in this population when compared with the findings of Ertekin-Taner *et al.* In that study, the authors selected pedigrees in which LOAD was known to segregate and which were identified via a proband with Aβ levels in the highest extreme. The study samples were therefore preselected for an increased probability of plasma Aβ heritability. Furthermore, in the Ertekin-Taner *et al.* study, plasma Aβ42 exhibited even higher heritability (~56%) [62, 204] than Aβ40. This is in contrast with the findings in the Vis study where significant heritability was not found for the plasma Aβ42 trait using the analysis methods implemented in SOLAR (h^2 = 4.6%). The Aβ42 peptide is generally considered to be more pathogenic than its shorter sibling; consequently in a study of LOAD pedigrees it is more likely to show significant heritability. However, in the context of the Vis population, the absence of significant heritability for this trait may indicate that plasma Aβ42 is not under significant genetic control. The implications of this are apparent in the genome-wide linkage analysis where analysis of Aβ42 did not identify any suggestive or significant linkage peaks. In contrast, analysis of the plasma Aβ40 trait did identify some interesting loci with suggestive linkage in the

Vis population. Encouragingly, three of these overlapped almost perfectly with three of eleven (27%) significant or suggestive linkage peaks reported in a meta-analysis of LOAD linkage studies [2]. This convergence, whilst not statistically significant, is noteworthy nonetheless and indicates that there may be one or more LOAD genes within these three linkage peaks that may function through an Aβ pathogenic pathway.

Overall the study did not identify any linkage peaks with a LOD score of greater than 3.3 (the generally accepted threshold for significance in a genome-wide linkage scan) for any of the four traits evaluated (Aβ40, Aβ42, Aβ40:42 ratio and Aβ40+42). The study had been designed to make use of techniques that increase power and reduce the effects of genetic heterogeneity, so the lack of significant linkage was disappointing. However the results of this study did confirm that plasma Aβ is a heritable quantitative trait even in populations that are not enriched for samples with cognitive deficits.

Given the convergent evidence for some of the suggestive linkage peaks reported in this study with LOAD linkage peaks, the question of which genes lay within the support interval of these peaks was investigated and these genes were then checked to see if any had been reported as candidate genes in LOAD association studies. To this end, a web resource within the Alzforum website called 'AlzGene' was used [31], which is reviewed in Chapter 1. Table 2.5 summarises, as others have done previously [2], the AlzGene genes within the support interval of the suggestive linkage peaks and it was encouraging to note that three AlzGene 'top genes'

(queried in March 2010) lay within the support intervals for three of the peaks. These are described in more detail below:

LOC645503 is a pseudogene which is described on the NCBI website as "similar to SUMO-1 activating enzyme subunit 2". SUMO is an acronym for **'S**mall **U**biquitin-like **Mo**difier', SUMO proteins bind to other proteins, as a form of post-translational modification, in a process commonly referred to as SUMOylation. There are many published studies reporting the effects of SUMOylation on different peptides. In 2007 Scheschonka *et al.* reviewed the role SUMOylation in relation to neuronal processes [415]. In 2003, Li *et al*. [416] published a report on the effect of SUMOylation on the production of Aβ and this has since been confirmed by others [417], implicating a role for SUMOylation in Alzheimer's disease pathogenesis. However, since *LOC645503* is a pseudogene, the convergent identification of association of this locus with AD and suggestive linkage with plasma Aβ may be indicative of a false positive.

CHRNB2: Variants in this gene have frequently been linked to epilepsy disorders and nicotine addiction [418-420] and there have also been several reports of association (both negative and positive) with Alzheimer's disease [232, 237, 324, 421-423]. For many years, the loss of cholinergic neurons has been studied in Alzheimer's disease and many of the drugs currently used to help manage the symptoms of AD function by inhibiting acetylcholinesterase (AChE) and thus increasing cholinergic activity (see above). The identification of a suggestive linkage peak to plasma Aβ located proximal to the *CHRNB2* locus is of interest given the possible functional interactions, which warrants additional investigation.

NEDD9: The NEDD9 protein has a SH3 domain and two SH2 domains and is believed play a role in focal adhesions, cell signalling and possibly apoptosis [424]. NEDD9 is expressed in brain, as its name 'neural precursor cell expressed, developmentally down-regulated 9' suggests, and several studies implicate a role for NEDD9 in development and differentiation of neuronal progenitor cells [425, 426]. A literature search did not reveal any evidence to suggest a direct role for *NEDD9* in modulating Aβ or related proteins and so it is likely that if variation at the *NEDD9* locus is responsible for linkage to Aβ its effect may be via an indirect or as yet undiscovered pathway or interaction.

Given that most genes reported in the AlzGene database have either been investigated for association with AD because of some *a priori* reason for candidacy, or have been identified through reports of a positive association with AD, these genes must represent interesting candidates and so their presence under a linkage peak for plasma Aβ40, albeit suggestive, will be of interest to the field. As illustrated in Table 2.5, the greatest abundance of 'AlzGene genes' was found within the linkage peaks on chromosome 1 that overlap with the LOAD linkage peaks reported by Butler *et al*. [2]. This may be due to the fact that many candidate gene studies have focused on reported LOAD linkage regions, so there is an over-representation of Alzgene genes at these loci.

GWAS analyses: Following on from the linkage analysis of plasma Aβ in subjects from Vis, a GWAS involving over 300,000 SNPs was carried out. An additional population from the neighbouring island of Korcula was also studied. The study

subjects recruited from the island of Korcula were likewise genotyped for over 300,000 SNPs and plasma concentrations of Aβ40 (Aβx-40) and Aβ42 (Aβx-42) in these subjects was also measured. The measurement of plasma Aβ40 in this second population differed from the measurement of this peptide in Vis due to a change in the capture antibody used and this should be considered when interpreting the results, in particular the meta-analysis of the Aβ40 trait. Encouragingly however, there was consistency between the two populations in the relationship between age and the plasma Aβ40 trait, indicating that the use of the different capture antibodies may not have affected the results significantly. The relationship between age and the Aβ40 trait has previously been reported by others [427]. Replication of this observation indicates that plasma Aβ40 has similar properties in these populations to those observed in outbred populations. This increases the possibility that the findings in this study will be of relevance to large urban populations.

The results of GWAS analysis of the four plasma Aβ traits in each population, both independently and using meta-analysis, whilst interesting and worthy of further investigation, was disappointing since it did not identify any SNPs that achieved genome-wide significant association following traditional Bonferroni correction for multiple tests. Several interesting loci were identified that met the criteria for suggestive association and these will be good candidates for follow-up in LOAD case control studies. This data will also be useful in helping to identify LOAD genes by careful and thoughtful use of the database of results in convergent analysis with other disease related traits, as illustrated in Chapter 3. It should also be noted that the use of the Bonferroni correction for multiple tests is likely to be over-conservative

for GWAS studies where the assumption of independent tests is violated by the presence of LD between SNPs. This is especially true for GWAS studies of isolated populations where larger regions of LD exist [269] within the genome and the probability of one or more markers being in LD is increased. For these reasons, the SNPs identified with the lowest p-values for each study are highlighted in Tables 2.7, 2.9 and 2.10. The SNPs and loci listed in these Tables should therefore be considered as providing strong evidence that they influence circulating Aβ concentrations and should be further evaluated in LOAD.

Finally, association was not observed with the *APOE* locus in either the Vis or Korcula studies and suggestive linkage to chromosome 19 was also not observed for Vis. This is not surprising given that previous studies evaluating plasma Aβ levels have also not identified significant association of this trait with the *APOE E4* genotype [204, 427]. This is in contrast to the highly significant association of this locus in AD samples. These observations collectively suggest that the molecular mechanisms by which the ApoE protein influences risk for AD do not act by influencing circulating levels of Aβ.

Trait heritability: When a polygenic model for the Vis traits was applied in SOLAR, significant heritability for the plasma Aβ40 trait was identified in this population ($p=0.041$) (Table 2.4). However, when a polygenic model was applied for the Vis traits in GenABEL, significant heritability was not found for any of the traits in this population, although the plasma Aβ42 trait revealed the strongest (but non-significant) evidence for heritability (Table 2.6). There are multiple possible

explanations for the differences observed in estimating heritability of these traits using the two methods. Primarily, it should be noted that the evidence for heritability of these traits in the Vis population is marginally significant using both SOLAR and GenABEL analyses. The sample subset used in each of the analyses was also slightly different, due to genotyping and quality control measures eliminating different subjects from each analysis. Furthermore, the kinship estimates used in the two methods are considerably different. In the SOLAR method, kinships reported by questionnaire and interviews were used to determine pedigrees. While this method is widely used, it is known to have problems, such as incorrect or incomplete reporting of familial relationships. The method used in the GenABEL analysis was not based on reported relationships but instead on a genomic kinship matrix which uses IBD values calculated between each pair of individuals based upon the genotypes gathered in the GWAS [318]. This method does have the problem of being subject to genotyping errors, but provided that careful and thorough genotype quality control is undertaken it is likely to produce an accurate portrayal of relationships in the population.

The analysis of plasma Aβ in subjects from the Korcula population revealed that, in this population, plasma Aβ concentrations are a heritable quantitative trait. All four plasma Aβ traits were significantly heritable, with plasma Aβ40 having a heritability of 46.7%, similar to the heritability values identified by Ertekin-Taner *et al*. in LOAD pedigrees [204]. [204]. In the Korcula study, plasma Aß40 was studied using BNT-77 as the capture antibody instead of BAN-50, which was used for capture in the Vis study. As discussed in Chapter 2, this change was made to improve the

accuracy of measurement by reducing non-specific signals (interference). The improved accuracy of measurement may well account for the increased heritability observed in the Korcula population, but the possibility that the two populations have a true difference in the heritability of plasma Aß40 cannot be eliminated. In this regard, it is important to emphasize that plasma Aß42, which showed significant heritability in the Korcula but not in the Vis population, was measured identically in the two populations.

Given the relatively high heritability of the plasma Aβ traits in the Korcula populations, it was surprising that none of the SNPs achieved genome-wide significant association in this population. This may indicate that the study was underpowered or that the trait is under the influence of multiple genetic factors with individually small effect sizes. A similar result has been observed with other traits such as height [428] and Alzheimer's disease [242, 243] where, despite high heritability (up to 80%), very large numbers of samples are needed to identify SNPs that achieve genome-wide significant association.

One of the aims of conducting a meta-analysis of the results from these two studies was to increase the sample size and thus increase the power to detect association. Unfortunately, the meta-analysis also failed to identify SNPs that achieved genome-wide significance and, with the exception of the plasma Aβ42 trait, all of the top hits for this analysis were in intergenic regions. The identification of SNPs with low p-values in intergenic regions for GWAS is a not an uncommon finding [429, 430]. Although these regions do not harbour genes there are many other alternative explanations for a true association at these loci. These include but are not limited to

the possibility that the SNP may lie within or be proximal to a microRNA sequence or may influence a long range enhancer or suppressor. It may also be possible that the meta-analysis was affected by the fact that Vis showed weak evidence of heritability, that the Aβ40 species measured in the two studies was slightly different, or that the study was still underpowered and needed additional samples or populations. Alternatively, it is possible that multiple SNPs in each population influence levels of plasma Aβ and that there is little similarity in the SNPs that play such a role between the populations. The study of these traits in additional populations will be important to both increase the power to detect association through meta-analysis and to provide replication of SNPs identified in this study. Furthermore, the results of this study can provide a useful tool for phenotype convergence analysis. To this effect, the top 19 SNPs identified in chapter two were evaluated for association with LOAD using the results from a published LOAD GWAS. This analysis did not identify any SNPs with a p-value less than 0.0026 which is the threshold for significance after correction for 19 tests. However there were 6 SNPs with an uncorrected p-value of less than 0.1, where by chance we would expect to see only 2.

These results in one respect indicate that the approach used in chapter 2 to identify likely LOAD candidate loci through identification of plasma Aβ loci may not have been successful. However it is known from various studies in the literature and is exampled in chapter 3 of this thesis, that a larger sample size is often required for identification of significant association with LOAD than for LOAD endophenotypes. Consequently the results presented in table 2.11 are inconclusive

given the relatively small sample size of the LOAD series (~730 cases and 1100 controls) and the SNP's will likely need to be genotyped in additional case-control samples. This is especially pertinent for the 6 SNPs identified in table 2.11 with a p-value of less than 0.1, as this over representation of SNPs with smaller p-values may be indicative of a true, but modest, association with LOAD for one or more of these SNPs that warrants further investigation.

In addition, three recently identified LOAD loci (*CLU*, *CR1* and *PICALM*), were investigated for association with plasma Aβ levels in the Vis and Korcula populations. The results did not indicate strong evidence for a role of these loci in influencing plasma Aβ levels in these populations where only the SNP proximal to the PICALM gene (rs3851179) showed nominally significant association with plasma Aβ. This finding is similar to that which has been observed for *APOE* where both this study and others have failed to identify association for this established LOAD risk locus with plasma Aβ levels. This result indicates that there may be molecular pathways that can influence risk for LOAD that do not have a direct impact on Aβ measured in plasma, the implications of which are open to interpretation and should be the subjects of further study.

5.4 Phenotypic convergence analysis identifies significant association of a functional IDE SNP.

As mentioned previously (Chapter 1), Aβ degrading enzymes are of particular importance in the study of Alzheimer's disease. Improved knowledge of enzymes that actively degrade Aβ has substantial implications for development of therapeutic

targeting strategies. The findings presented in Chapter 3 provide compelling evidence that variants at the *IDE* locus can influence AD through altered *IDE* expression and subsequently altered Aβ levels.

The approach implemented here was to first identify SNPs that are likely to be functional by selecting SNPs located in conserved regions. These SNPs were then evaluated for association with the expression levels of the target candidate gene *IDE*. Since the SNPs were selected based upon conservation, it is likely that SNPs that then show association with altered expression levels may directly represent a functional SNP. In this case, the target gene was a known Aβ degrading enzyme (*IDE*). Consequently the effect of the identified eSNP(s) could be directly tested for association with altered levels of Aβ *in vivo* by evaluating a proxy (rs7910977) for the eSNP (rs6583817) that was genotyped in the isolated populations of Vis and Korcula. A SNP that influences expression of *IDE* may also have the potential to influence plasma Aβ concentration.

The result indicated that plasma Aβ did indeed associate with the eSNP proxy and in such a way that the allele that correlated with increased *IDE* expression likewise correlated with decreased plasma Aβ in both populations. The third stage of the convergence analysis was to determine if this SNP(s) could also influence AD risk. Indeed, analysis of rs6583817 indicated association of the minor allele with decreased risk for LOAD which is consistent with the previous findings. Furthermore, evaluation of rs6583817 in functional test models supports the belief that this is in fact a functional SNP that has a direct effect on *IDE* expression.

The results presented here strongly support a role for *IDE* in LOAD pathogenesis, however the data indicate that the effect of altered IDE expression on risk for LOAD is small (OR = 0.87). It is unlikely that the potentially functional variant identified in this study would have been identified using GWAS methods that predominantly identify common variants with larger effect sizes than the SNP identified here. Thus this phenotype convergence approach represents a relevant avenue of research for future studies.

As discussed in Chapter 1, the effect size of identified LOAD genes, with the exception of *APOE*, is consistently weak. The implications of this for future use of identified genetic risk loci as tools for predicting risk of disease in individuals thus becomes challenging and at this time is unlikely to be feasible. Furthermore, therapeutic targeting of biological candidates that do not show a substantial influence on disease risk is also less appealing than when there is little evidence for a substantial influence on disease. The data presented here however evaluate the effect of a human *IDE* SNP that influences expression of the associated gene on AD risk. If IDE or other Aβ degrading enzymes were to be targeted for therapeutic intervention, alteration of the expression or activity of the targeted enzyme would probably have a far greater affect than the SNP effect and over the normal range; so potentially providing a useful therapeutically effect. Thus, the identification of genetic variants that influence risk for LOAD, even with small effect sizes is, of potential importance for future research into this disease.

5.5 Evaluation of approaches for studying mitochondrial DNA damage in disease.

The emergence of a role for mitochondrial dysfunction in Alzheimer's disease and subsequent work indicating that Aβ can interact with mitochondrial proteins represents a relatively novel molecular mechanism influencing AD. This has led to a renewed interest in understanding the role that this organelle plays in AD and other neurodegenerative diseases. This avenue of research has also led to the development of multiple hypotheses to account for the involvement of mitochondria in AD.

There are many possibilities for testing mitochondrial hypotheses in LOAD, some of which include:

1. Investigation of mitochondrial haplogroups or homoplasmic variants which can be tested for association with disease. Like most nuclear-encoded SNPs tested in candidate genes, the reports of this type of association are mixed. Similarly, the common 4997bp mtDNA deletion can be tested for association with disease. Published association studies of common mitochondrial variants in AD have recently been added to the AlzGene database[31]

2. Study of mitochondrial mutational burden. Measurement of the cumulative number of rare, heteroplasmic variants in the mitochondrial genome as a whole or in individual genes or groups of genes in order to test the hypothesis that AD patients have a higher mutational burden than control subjects.

3. Measurement of mtDNA damage and its association with disease. Here, a general measure of deletions, cross-links, adducts and other mtDNA modifications that are detectable by the qPCR assay which have the common property of inhibiting processivity of the polymerase and hence PCR efficiency.

4. Candidate nuclear genes can be tested for association with disease or possibly for association with a marker for oxidative stress or mitochondrial dysfunction (mitochondrial quantitative trait).

In this thesis, mitochondrial mutational burden and mtDNA damage were explored for association with age and neurodegenerative disease. Additionally a candidate gene study was conducted evaluating SNPs for association with LOAD in nuclear genes that encode essential components of the mitochondrial polymerase γ and electron transport chain complex IV.

Mitochondrial mutational burden: Others have previously reported an increase in mutational burden with age. In 2002 Lin *et al* [295] reported an evaluation of mitochondrial mutational burden using mtDNA isolated from brain tissue. The authors identified a mutational burden of 2.89E-04 mutations/bp (or 289/million bp) in elderly control subjects and 1.21E-04 (or 121 per million bp) in young subjects [431]. However, this age related analysis compared the mitochondrial mutational burden of only two data sets, young (<31 yrs old) and elderly (>53 yrs old) with a sample size of 14 and 10 respectively. The data described in Chapter 4, in contrast, investigated four age groups (sample size of 3 per group) and suggests a more

complex relationship between age and mitochondrial mutational burden, at least with respect to mtDNA isolated from blood. The mutational burden estimates reported by Lin *et al.* [295] also indicate a higher mutational burden in mtDNA isolated from brain tissue than that from blood. This is highlighted by the fact that in the study described in chapter 4, the *HVR* locus, which is expected to have a higher mutational burden than other loci in the mt genome [280], has a lower mutational burden than the *COI* locus studied by Lin *et al*. Despite the higher mutational burden reported in the brain, the values are generally remarkably similar in magnitude (22-160/million bp vs 121-289/million bp). Furthermore the cell turnover of leukocytes is much higher than brain cells since leukocytes are derived from erythropoietic stem cells, which may be under strong selection pressure to eliminate cells that have accumulated substantial mtDNA damage or mutation. The opportunity for leukocytes to survive in the presence of acquired somatic mutations is therefore probably substantially less than that for brain cells, so it is not surprising that the mitochondrial burden is higher in the brain.

The mitochondrial mutational burden estimates in chapter 4 were obtained by using a PCR-cloning-sequencing approach the utility of which has been previously demonstrated by others [287, 385-387, 432]. However since these experiments were completed, alternative sequencing methods have been developed that are capable of evaluating the entire mitochondrial genome instead of 1kb targets by use of new technology collectively referred to as ‘next-generation’ (next-gen) sequencing. At the time of the experiments presented in chapter 4, next-gen sequencing was in its infancy and was not a viable technology for most studies due to economic

constraints. Furthermore the ability of this technology to detect extremely rare somatic mutations (heteroplasmy) in a sample comprised of many mitochondrial genomes was controversial. Improvements have been made to this technology which may improve the efficacy of this approach in the future, however current error rates [433] and the reduced capacity for detecting very rare variants maintains the relevance of the PCR-cloning-sequencing approach, used here, at this time. It may be beneficial in the future to combine the next-gen sequencing approaches to identify common mutations and to estimate heteroplasmy [434], followed by the targeted PCR-cloning-sequencing approach to confirm mutational burden estimates for very rare variants,

A recent paper that evaluated the relationship between oxidative stress and mitochondrial DNA damage or mutation, concluded that despite extensive ROS damage to mtDNA, acquired mutations in mtDNA are kept at low levels [432]. It was suggested by the authors that this might result from a mechanism by which damaged DNA molecules are either repaired by base excision repair or, DNA molecules are linearised and degraded by unknown exonucleases. Unrepaired pre-mutagenic base lesions, such as 8-hydroxyguanosine (8-oxo-G), lead to mutations (mostly transversions). Surprisingly, oxidative damage (by hydrogen peroxide) caused ten times as many single strand breaks (SSB) and apurinic sites compared with oxidized pre-mutagenic base changes. The degradation mechanism was found to be unique to mtDNA, presumably because there are multiple copies of mtDNA but not nDNA. It suggests that while mtDNA mutations may not increase

dramatically with age (because they are either repaired or degraded) degradation of oxidatively damaged mtDNA may increase with age, leading to low copy number. Evaluating the extent of DNA damage by qPCR or copy number assay, rather than mutational burden, may therefore be more reflective of the overall level of ROS induced damage or functional impairment of mtDNA. This was explored in this thesis for mtDNA isolated from the frozen brain tissue of patients diagnosed with various neurodegenerative diseases and controls.

Mitochondrial DNA damage: DNA damage can be assessed using qPCR methods based upon the hypothesis that molecular DNA lesions such as adducts, base modifications (e.g. 8oxoG) and strand breaks will result in stalling of the PCR polymerase in a way that is measurable [435, 436]. Previously, others have used qPCR to measure DNA damage by estimating mitochondrial copy number using short qPCR assays and then measuring damage by long qPCR, with the difference in copies detected by both methods serving as an indicator of the damage. This is reviewed by Senyene Hunter *et al*. 2010 [437]. The study presented in this thesis however has demonstrated that DNA damage can be detected even in short PCR amplicons. Consequently it is possible that estimating the mitochondrial copy number using short amplicons may result in confounding copy number change with DNA damage. Both the study presented in Chapter 4 and the Mambo *et al*. study [386] support this observation.

An alternative way to use qPCR to evaluate DNA damage is to use Sybr green. The advantage of this is that larger target regions can be investigated as the technique

does not rely on annealing of a probe to the target DNA sequence. However Sybr green is known to cause some inhibition of the PCR reaction which may also be a potential confounder [438]. Additionally biochemical assays are available that allow for the measurement of 8-oxoguanine (8oxoG), a molecular lesion that is both highly mutagenic and believed to be the most frequent lesion associated with oxidative DNA damage. However these approaches limit the study to evaluation of just one oxidative DNA lesion and as such do not provide the same breadth of analysis as the qPCR approach. In summary there are many ways of measuring mitochondrial DNA damage; all methods have caveats and potential confounders that must be controlled. Consequently, whilst the method employed here relies on certain assumptions and also has possible confounders, the alternatives do not provide an obviously better solution.

The data presented in this thesis evaluated mitochondrial DNA damage in samples of frozen brain tissue isolated from the temporal cortex (frequently affected by LOAD pathology) and the cerebellum (rarely affected by LOAD pathology). The level of mtDNA damage was investigated in these two brain regions for four groups of elderly subjects defined by pathological criteria into three disease categories and a normal control group. mtDNA damage was measured using qPCR techniques and the analysis approach outlined in Chapter 4 allowed for evaluation of mtDNA damage independent of estimation of mitochondrial DNA copy number.

The results of this study imply an association between neurodegenerative diseases and mitochondrial DNA damage which is especially high for samples isolated from the brains of subjects diagnosed with VaD. Interestingly in the case of the vascular

dementia (VaD) samples, there was no significant difference in the levels of DNA damage between the two areas of the brain sampled: Cerebellum and temporal cortex (Table 4.8). This result indicates that the brain tissue of VaD patients may undergo high levels of non-selective oxidative DNA damage either during the course of the disease process or possibly upon the death of the patient as a result of a fatal stroke. The results observed for the VaD subjects warrants further investigation in additional samples and may present informative novel data for VaD research.

The results from the analysis of the PSP and normal subjects (Table 4.8) both indicated a significant increase in mtDNA damage in the DNA isolated from the temporal cortex when compared with the cerebellum at all three mitochondrial loci investigated. The results also indicate that whilst the temporal cortex exhibited more damage than the cerebellum for PSP subjects, there was reduced damage at the temporal cortex when compared with controls. This indicates that increased DNA damage at the temporal cortex may occur as a function of normal aging without associated cognitive deficits or pathology. Furthermore the decreased levels of damage observed in the temporal cortex for PSP subjects could represent either a protective effect against oxidative damage in these subjects or it is possible that damaged mtDNA molecules in these subjects has been selectively degraded.

When mtDNA damage observed for AD subjects was compared with that of controls (Table 4.8) it was surprising to note that there were no significant differences between these two sets of subjects for any of the mitochondrial loci examined or the two tissues sampled. It may be possible that the results presented here indicate that indeed mtDNA damage is not a significant factor in AD pathogenesis or, as

proposed above for the PSP subjects, damaged molecules are selectively degraded. This would indicate that the majority of DNA molecules that are intact are as yet not damaged beyond that observed in normal aging. However, the data presented in Table 4.8 did identify a significant increase in the levels of mtDNA damage at the *COII* locus for mtDNA isolated from the temporal cortex of AD subjects, when compared with that isolated from the cerebellum. This finding is of interest since there have been multiple reports of cytochrome oxidase deficiency in LOAD samples [303, 439, 440] and the *COII* gene encodes one of the three catalytic subunits of this OXPHOS enzyme.

Overall the 16s locus was the most damaged locus in normal control subjects and could indicate a higher tolerance for DNA damage at this locus when compared with the D-loop and COII loci. As suggested by Shokolenko *et al*, the mitochondria may degrade damaged and mutated mitochondrial DNA molecules. It is not beyong the realm of possibility that mtDNA molecules with damage detected at certain loci of the mitochondrial genome are more preferentially degraded and consequently mtDNA molecules with damage at the COII and D-loop loci have been degraded more actively than those molecules with damage at the 16s locus. This effect was noted by Mambo et al [386] where the 'repair time' following damage of mtDNA with chemical agents varied across the different loci evaluated.

The interpretation of the data presented here has to be considered in the context of a tissue homogenate which means that the proportion of neuronal cells in each homogenate is unknown and likely to be variable across different subjects, diseases and locations in the brains from which the samples were isolated. It would be

pertinent to follow this study up in a larger number of samples with mtDNA isolated from neuronal cells only (e.g. by laser capture microdissection) to determine if this data is reflective of neuronal cells or perhaps other brain cells such as glial cells. Furthermore, the sample size used in this study is relatively small (10 subjects per group), so additional studies would benefit from a larger sample size, although this may be limited by the availability of the appropriate tissue. Exploration of additional loci or optimisation of long-range PCR techniques may improve the investigation of mtDNA damage in future studies.

Candidate gene association study: The mitochondrial genome encodes essential proteins involved in mitochondrial function and as such can have important implications in disease, especially when mitochondrial dysfunction is implicated. However the majority of the proteins that are present within the mitochondria and are likewise important to its function are encoded by the nuclear genome [383]. One example of a mitochondrial disease that can be caused by mutations in the nuclear genome is progressive external opththalmoplegia (PEO). In this disease mutations in the genes that form the components of the mitochondrial DNA polymerase (*POLG*, *POLG2*, *PEO1*) cause impaired mitochondrial replication and repair resulting in reduced mtDNA copy number and the accumulation of mtDNA deletions and mutations[441-443]. The onset of symptoms of this disease can range from 18 to 40 years [444] indicating that the effect of impaired mitochondrial replication and proof reading activities requires a significant amount of time to progress prior to the onset of symptoms. It could therefore be possible that late-onset disease such as AD and

PD might also be caused by defects in the mitochondrial polymerase where longer periods of time are needed before the effects of loss of mitochondrial DNA becomes apparent. Consequently the three genes that form the principle components of the mitochondrial polymerase gamma were investigated for SNPs that show association with LOAD. Interestingly two SNPs included in this study showed nominal significant association with LOAD (Table 4.9) although neither maintained significance following bonferonni correction for multiple tests. However given the logical rationale behind the pursuit of SNPs at these loci for association with LOAD, genotyping of additional samples to increase power for detection of SNPs with small effect size might confirm the association detected here.

In addition to the three genes described previously, a candidate gene study was also carried out to investigate SNPs in 17 nuclear genes that encode components of cytochrome oxidase (complex IV) of the mitochondrial electron transport chain. The rationale behind the selection of these genes was that several studies have reported decreased activity of complex IV in brain and platelet samples for AD patients [302, 303, 390, 439]. One cause of this observed decreased activity could be impaired function or expression of the structural components of this complex, all of which are encoded by nuclear genes. Association analysis of SNPs at these 17 loci identified 70 SNPs within each of the genes and their 20kb flanks, of which four exhibited nominally significant association with LOAD, localized to two genes. Interestingly the two genes identified in this study encode two isoforms of the same subunit of complex IV, namely the COX6B subunit. Both of these genes are located on chromosome 19, however they are located approximately 20 million base pairs apart

and 10 million bps from the *APOE* locus indicating independent effects. Whilst none of the four SNPs identified at the locus for these 2 genes maintain significance following bonferonni correction for multiple tests, they do represent a novel finding and warrant further evaluation in additional samples.

The limitations of the candidate gene study are firstly the narrow age range of the subjects studied - aged 60-80 years. If there are SNPs that interact with age such that risk for AD is only realized in older samples then these would be missed. Secondly, the size of the study (788 AD and 1165 control samples) may be too small to identify SNPs with small effect sizes. This possibility was evident in the recent LOAD GWAS studies, which contained sample sizes of many thousands of subjects to reliably identify LOAD risk SNPs. On the other hand, this study provides insight into complex IV genes that may play a role in LOAD aetiology and follow-up studies using a subset of the SNPs identified here would permit a more extensive evaluation of these genes.

5.6 Conclusions

The aim of the projects presented in this thesis was to identify novel genetic risk factors for LOAD.

Evaluation of plasma Aβ concentrations in two isolated populations identified significant heritability of plasma Aβ40, however the evidence for Aβ42 was much less convincing. This result indicates that genetic factors may not explain as much of the variation in plasma Aβ concentrations (especially Aβ42) in populations that are not selected for cognitive deficits than those that are. Although these studies did not

identify significant linkage or association of genetic variation with the plasma Aβ traits tested, several promising candidate genes were identified. Furthermore the data generated from these studies is a valuable resource for phenotype convergence analysis as exampled for the *IDE* locus in chapter 3.

Finally, exploration of mitochondrial genetic risk factors for LOAD surprisingly identified high levels of mtDNA damage in VaD patients; however the evidence for AD was less compelling. These studies also highlighted the need for careful consideration of methods and controls to prevent amplification of *numts* or *in vitro* DNA damage that can confound the interpretation of the results. Furthermore the evaluation of nuclear encoded proteins essential for mitochondrial function may be an overlooked area of research into the genetics of LOAD. The identification of novel candidate LOAD genes in this study that play a role in mitochondrial function warrant further evaluation.

5.7 Future directions

In an extension of the plasma Aβ GWAS analysis, the R package ProbABEL, can be used to facilitate imputation of additional genotypes in the two populations studies in this thesis. Analysis of imputed SNP data will enable the evaluation of additional SNPs in each population that may not have been 'tagged' by the more common variants genotyped in each study. Any SNP association identified using imputed data will need to be validated by genotyping the samples for the SNPS in question using traditional genotyping techniques such as Taqman.

The evaluation of additional populations for GWAS with plasma Aβ is another plausible approach for both evaluating replication of SNPs of interest identified in the first two studies and for increasing power in subsequent meta-analysis.

Additionally measurement of plasma Aβ40 levels in Vis using the BNT77 capture antibody as used for the Korcula study might also improve the meta analysis for this trait.

Studies of nuclear genetics in LOAD have, until recently, failed to unequivocally identify variation that is associated with the disease beyond the well known *APOE E4* association. It was only the use of GWAS in very larger numbers of cases and controls that advanced the field to the point of being able to identify SNPs that replicate reliably. This could likewise be the case with common mtDNA variations and perhaps one avenue of research that should be pursued is the evaluation of common mtDNA variants and haplogroups in very large case control series. The latter can be achieved most easily by high throughput sequencing of the relatively small mitochondrial genome in many thousands of cases and controls.

Finally the possibility that damaged mtDNA molecules, that cannot be repaired, are degraded opens up a new avenue of research into the complexities of the relationship between mitochondrial DNA damage and disease. To this end the expression levels of mtDNA degrading enzymes could be evaluated to see if they are upregulated in AD. Exploration of the mechanisms that lead to identification and degradation of damaged mtDNA molecules may implicate additional molecular pathways that can be manipulated therapeutically.

6. References

1. Ruiz-Pesini, E., et al., *An enhanced MITOMAP with a global mtDNA mutational phylogeny.* Nucleic Acids Res, 2007. **35**(Database issue): p. D823-8.
2. Butler, A.W., et al., *Meta-analysis of linkage studies for Alzheimer's disease--a web resource.* Neurobiol Aging, 2009. **30**(7): p. 1037-47.
3. Carrasquillo, M.M., et al., *Concordant association of insulin degrading enzyme gene (IDE) variants with IDE mRNA, Abeta, and Alzheimer's disease.* PLoS One, 2010. **5**(1): p. e8764.
4. Alzheimer, A., *U¨ ber eine eigenartige Erkrankung der Hirnrinde.* Allg. Z. Psychiat. Psych.-Gerichtl. Med., 1907. **64**: p. 146-148.
5. Alzheimer, A., et al., *An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde".* Clin Anat, 1995. **8**(6): p. 429-31.
6. Alzheimer, A., *U¨ ber eigenartige Krankheitsfa¨lle des spa¨teren Alters.* Zbl. ges. Neurol. Psych., 1911. **4**: p. 356-385.
7. Alzheimer, A., H. Forstl, and R. Levy, *On certain peculiar diseases of old age.* Hist Psychiatry, 1991. **2**(5 Pt 1): p. 71-101.
8. Tomlinson, B.E., G. Blessed, and M. Roth, *Observations on the brains of demented old people.* J Neurol Sci, 1970. **11**(3): p. 205-42.
9. Kukull, W.A., et al., *Dementia and Alzheimer disease incidence: a prospective cohort study.* Arch Neurol, 2002. **59**(11): p. 1737-46.
10. Ott, A., et al., *Prevalence of Alzheimer's disease and vascular dementia: association with education. The Rotterdam study.* BMJ, 1995. **310**(6985): p. 970-3.
11. Tierney, M.C., et al., *The NINCDS-ADRDA Work Group criteria for the clinical diagnosis of probable Alzheimer's disease: a clinicopathologic study of 57 cases.* Neurology, 1988. **38**(3): p. 359-64.
12. Braak, H. and E. Braak, *Neuropathological stageing of Alzheimer-related changes.* Acta Neuropathol, 1991. **82**(4): p. 239-59.
13. Hyman, B.T. and J.Q. Trojanowski, *Consensus recommendations for the postmortem diagnosis of Alzheimer disease from the National Institute on Aging and the Reagan Institute Working Group on diagnostic criteria for the neuropathological assessment of Alzheimer disease.* J Neuropathol Exp Neurol, 1997. **56**(10): p. 1095-7.
14. Mirra, S.S., et al., *The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease.* Neurology, 1991. **41**(4): p. 479-86.
15. Jellinger, K.A., *Clinicopathological analysis of dementia disorders in the elderly--an update.* J Alzheimers Dis, 2006. **9**(3 Suppl): p. 61-70.
16. Petersen, R.C., et al., *Aging, memory, and mild cognitive impairment.* Int Psychogeriatr, 1997. **9 Suppl 1**: p. 65-9.
17. Petersen, R.C., et al., *Mild cognitive impairment: clinical characterization and outcome.* Arch Neurol, 1999. **56**(3): p. 303-8.

18. Petersen, R.C., *Mild cognitive impairment as a diagnostic entity.* J Intern Med, 2004. **256**(3): p. 183-94.
19. Klunk, W.E., et al., *Imaging the pathology of Alzheimer's disease: amyloid-imaging with positron emission tomography.* Neuroimaging Clin N Am, 2003. **13**(4): p. 781-9, ix.
20. Klunk, W.E., et al., *Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B.* Ann Neurol, 2004. **55**(3): p. 306-19.
21. Klunk, W.E., et al., *The binding of 2-(4'-methylaminophenyl)benzothiazole to postmortem brain homogenates is dominated by the amyloid component.* J Neurosci, 2003. **23**(6): p. 2086-92.
22. Grundke-Iqbal, I., et al., *Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology.* Proc Natl Acad Sci U S A, 1986. **83**(13): p. 4913-7.
23. Iqbal, K., et al., *Defective brain microtubule assembly in Alzheimer's disease.* Lancet, 1986. **2**(8504): p. 421-6.
24. Kidd, M., *Paired helical filaments in electron microscopy of Alzheimer's disease.* Nature, 1963. **197**: p. 192-3.
25. Terry, R.D., *The Fine Structure of Neurofibrillary Tangles in Alzheimer's Disease.* J Neuropathol Exp Neurol, 1963. **22**: p. 629-42.
26. Grundke-Iqbal, I., et al., *Microtubule-associated protein tau. A component of Alzheimer paired helical filaments.* J Biol Chem, 1986. **261**(13): p. 6084-9.
27. Lee, V.M., et al., *A68: a major subunit of paired helical filaments and derivatized forms of normal Tau.* Science, 1991. **251**(4994): p. 675-8.
28. Alonso, A., et al., *Hyperphosphorylation induces self-assembly of tau into tangles of paired helical filaments/straight filaments.* Proc Natl Acad Sci U S A, 2001. **98**(12): p. 6923-8.
29. Bancher, C., et al., *Accumulation of abnormally phosphorylated tau precedes the formation of neurofibrillary tangles in Alzheimer's disease.* Brain Res, 1989. **477**(1-2): p. 90-9.
30. Hutton, M., et al., *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.* Nature, 1998. **393**(6686): p. 702-5.
31. Bertram, L., et al., *Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database.* Nat Genet, 2007. **39**(1): p. 17-23.
32. McGeer, P.L., J. Rogers, and E.G. McGeer, *Inflammation, anti-inflammatory agents and Alzheimer disease: the last 12 years.* J Alzheimers Dis, 2006. **9**(3 Suppl): p. 271-6.
33. Glenner, G.G. and C.W. Wong, *Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein.* Biochem Biophys Res Commun, 1984. **120**(3): p. 885-90.
34. Masters, C.L., et al., *Amyloid plaque core protein in Alzheimer disease and Down syndrome.* Proc Natl Acad Sci U S A, 1985. **82**(12): p. 4245-9.
35. Kang, J., et al., *The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor.* Nature, 1987. **325**(6106): p. 733-6.
36. Joachim, C.L., J.H. Morris, and D.J. Selkoe, *Diffuse senile plaques occur commonly in the cerebellum in Alzheimer's disease.* Am J Pathol, 1989. **135**(2): p. 309-19.

37. Braak, H., et al., *Alzheimer's disease: amyloid plaques in the cerebellum.* J Neurol Sci, 1989. **93**(2-3): p. 277-87.
38. Friedrich, R.P., et al., *Mechanism of amyloid plaque formation suggests an intracellular basis of Abeta pathogenicity.* Proc Natl Acad Sci U S A, 2010. **107**(5): p. 1942-7.
39. Harper, J.D. and P.T. Lansbury, Jr., *Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins.* Annu Rev Biochem, 1997. **66**: p. 385-407.
40. Dickson, D.W., et al., *Correlations of synaptic and pathological markers with cognition of the elderly.* Neurobiol Aging, 1995. **16**(3): p. 285-98; discussion 298-304.
41. Katzman, R., *Alzheimer's disease.* N Engl J Med, 1986. **314**(15): p. 964-73.
42. Terry, R.D., et al., *Physical basis of cognitive alterations in Alzheimer's disease: synapse loss is the major correlate of cognitive impairment.* Ann Neurol, 1991. **30**(4): p. 572-80.
43. Braak, H., et al., *Vulnerability of cortical neurons to Alzheimer's and Parkinson's diseases.* J Alzheimers Dis, 2006. **9**(3 Suppl): p. 35-44.
44. Atwood, C.S., et al., *Senile plaque composition and posttranslational modification of amyloid-beta peptide and associated proteins.* Peptides, 2002. **23**(7): p. 1343-50.
45. Jack, C.R., Jr., et al., *Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia.* Neurology, 2002. **58**(5): p. 750-7.
46. Jack, C.R., Jr., et al., *Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade.* Lancet Neurol, 2010. **9**(1): p. 119-28.
47. St George-Hyslop, P.H., et al., *The genetic defect causing familial Alzheimer's disease maps on chromosome 21.* Science, 1987. **235**(4791): p. 885-90.
48. Tanzi, R.E., et al., *Amyloid beta protein gene: cDNA, mRNA distribution, and genetic linkage near the Alzheimer locus.* Science, 1987. **235**(4791): p. 880-4.
49. Goldgaber, D., et al., *Characterization and chromosomal localization of a cDNA encoding brain amyloid of Alzheimer's disease.* Science, 1987. **235**(4791): p. 877-80.
50. Tanzi, R.E., et al., *Protease inhibitor domain encoded by an amyloid protein precursor mRNA associated with Alzheimer's disease.* Nature, 1988. **331**(6156): p. 528-30.
51. Ponte, P., et al., *A new A4 amyloid mRNA contains a domain homologous to serine proteinase inhibitors.* Nature, 1988. **331**(6156): p. 525-7.
52. Kitaguchi, N., et al., *Novel precursor of Alzheimer's disease amyloid protein shows protease inhibitory activity.* Nature, 1988. **331**(6156): p. 530-2.
53. Ropper, A.H. and R.S. Williams, *Relationship between plaques, tangles, and dementia in Down syndrome.* Neurology, 1980. **30**(6): p. 639-44.
54. Mann, D.M., P.O. Yates, and B. Marcyniuk, *Some morphometric observations on the cerebral cortex and hippocampus in presenile*

Alzheimer's disease, senile dementia of Alzheimer type and Down's syndrome in middle age. J Neurol Sci, 1985. **69**(3): p. 139-59.
55. Busciglio, J., et al., *Generation of beta-amyloid in the secretory pathway in neuronal and nonneuronal cells.* Proc Natl Acad Sci U S A, 1993. **90**(5): p. 2092-6.
56. Haass, C., et al., *Amyloid beta-peptide is produced by cultured cells during normal metabolism.* Nature, 1992. **359**(6393): p. 322-5.
57. Seubert, P., et al., *Isolation and quantification of soluble Alzheimer's beta-peptide from biological fluids.* Nature, 1992. **359**(6393): p. 325-7.
58. Shoji, M., et al., *Production of the Alzheimer amyloid beta protein by normal proteolytic processing.* Science, 1992. **258**(5079): p. 126-9.
59. Vigo-Pelfrey, C., et al., *Characterization of beta-amyloid peptide from human cerebrospinal fluid.* J Neurochem, 1993. **61**(5): p. 1965-8.
60. Dovey, H.F., et al., *Cells with a familial Alzheimer's disease mutation produce authentic beta-peptide.* Neuroreport, 1993. **4**(8): p. 1039-42.
61. Nakamura, T., et al., *Amyloid beta protein levels in cerebrospinal fluid are elevated in early-onset Alzheimer's disease.* Ann Neurol, 1994. **36**(6): p. 903-11.
62. Ertekin-Taner, N., et al., *Linkage of plasma Abeta42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees.* Science, 2000. **290**(5500): p. 2303-4.
63. Burdick, D., et al., *Assembly and aggregation properties of synthetic Alzheimer's A4/beta amyloid peptide analogs.* J Biol Chem, 1992. **267**(1): p. 546-54.
64. Hilbich, C., et al., *Aggregation and secondary structure of synthetic amyloid beta A4 peptides of Alzheimer's disease.* J Mol Biol, 1991. **218**(1): p. 149-63.
65. Jarrett, J.T., E.P. Berger, and P.T. Lansbury, Jr., *The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease.* Biochemistry, 1993. **32**(18): p. 4693-7.
66. Gravina, S.A., et al., *Amyloid beta protein (A beta) in Alzheimer's disease brain. Biochemical and immunocytochemical analysis with antibodies specific for forms ending at A beta 40 or A beta 42(43).* J Biol Chem, 1995. **270**(13): p. 7013-6.
67. Walsh, D.M. and D.J. Selkoe, *A beta oligomers - a decade of discovery.* J Neurochem, 2007. **101**(5): p. 1172-84.
68. Lue, L.F., et al., *Soluble amyloid beta peptide concentration as a predictor of synaptic change in Alzheimer's disease.* Am J Pathol, 1999. **155**(3): p. 853-62.
69. McLean, C.A., et al., *Soluble pool of Abeta amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease.* Ann Neurol, 1999. **46**(6): p. 860-6.
70. Wang, J., et al., *The levels of soluble versus insoluble brain Abeta distinguish Alzheimer's disease from normal and pathologic aging.* Exp Neurol, 1999. **158**(2): p. 328-37.

71. Walsh, D.M., et al., *Amyloid-beta oligomers: their production, toxicity and therapeutic inhibition.* Biochem Soc Trans, 2002. **30**(4): p. 552-7.
72. Mohajeri, M.H., et al., *Passive immunization against beta-amyloid peptide protects central nervous system (CNS) neurons from increased vulnerability associated with an Alzheimer's disease-causing mutation.* J Biol Chem, 2002. **277**(36): p. 33012-7.
73. Chartier-Harlin, M.C., et al., *Early-onset Alzheimer's disease caused by mutations at codon 717 of the beta-amyloid precursor protein gene.* Nature, 1991. **353**(6347): p. 844-6.
74. Goate, A., et al., *Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease.* Nature, 1991. **349**(6311): p. 704-6.
75. Murrell, J., et al., *A mutation in the amyloid precursor protein associated with hereditary Alzheimer's disease.* Science, 1991. **254**(5028): p. 97-9.
76. Naruse, S., et al., *Mis-sense mutation Val----Ile in exon 17 of amyloid precursor protein gene in Japanese familial Alzheimer's disease.* Lancet, 1991. **337**(8747): p. 978-9.
77. Mullan, M., et al., *A pathogenic mutation for probable Alzheimer's disease in the APP gene at the N-terminus of beta-amyloid.* Nat Genet, 1992. **1**(5): p. 345-7.
78. Di Fede, G., et al., *A recessive mutation in the APP gene with dominant-negative effect on amyloidogenesis.* Science, 2009. **323**(5920): p. 1473-7.
79. Tomiyama, T., et al., *A new amyloid beta variant favoring oligomerization in Alzheimer's-type dementia.* Ann Neurol, 2008. **63**(3): p. 377-87.
80. Rovelet-Lecrux, A., et al., *APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy.* Nat Genet, 2006. **38**(1): p. 24-6.
81. Sleegers, K., et al., *APP duplication is sufficient to cause early onset Alzheimer's dementia with cerebral amyloid angiopathy.* Brain, 2006. **129**(Pt 11): p. 2977-83.
82. Schellenberg, G.D., et al., *Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14.* Science, 1992. **258**(5082): p. 668-71.
83. Van Broeckhoven, C., et al., *Mapping of a gene predisposing to early-onset Alzheimer's disease to chromosome 14q24.3.* Nat Genet, 1992. **2**(4): p. 335-9.
84. Mullan, M., et al., *A locus for familial early-onset Alzheimer's disease on the long arm of chromosome 14, proximal to the alpha 1-antichymotrypsin gene.* Nat Genet, 1992. **2**(4): p. 340-2.
85. Sherrington, R., et al., *Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease.* Nature, 1995. **375**(6534): p. 754-60.
86. Cruts, M. and C. Van Broeckhoven, *Molecular genetics of Alzheimer's disease.* Ann Med, 1998. **30**(6): p. 560-5.
87. Levy-Lahad, E., et al., *Candidate gene for the chromosome 1 familial Alzheimer's disease locus.* Science, 1995. **269**(5226): p. 973-7.

88. Rogaev, E.I., et al., *Familial Alzheimer's disease in kindreds with missense mutations in a gene on chromosome 1 related to the Alzheimer's disease type 3 gene.* Nature, 1995. **376**(6543): p. 775-8.
89. De Strooper, B., *Aph-1, Pen-2, and Nicastrin with Presenilin generate an active gamma-Secretase complex.* Neuron, 2003. **38**(1): p. 9-12.
90. Kimberly, W.T., et al., *The transmembrane aspartates in presenilin 1 and 2 are obligatory for gamma-secretase activity and amyloid beta-protein generation.* J Biol Chem, 2000. **275**(5): p. 3173-8.
91. De Strooper, B., et al., *Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein.* Nature, 1998. **391**(6665): p. 387-90.
92. Cai, X.D., T.E. Golde, and S.G. Younkin, *Release of excess amyloid beta protein from a mutant amyloid beta protein precursor.* Science, 1993. **259**(5094): p. 514-6.
93. Citron, M., et al., *Mutation of the beta-amyloid precursor protein in familial Alzheimer's disease increases beta-protein production.* Nature, 1992. **360**(6405): p. 672-4.
94. Suzuki, N., et al., *An increased percentage of long amyloid beta protein secreted by familial amyloid beta protein precursor (beta APP717) mutants.* Science, 1994. **264**(5163): p. 1336-40.
95. Scheuner, D., et al., *Secreted amyloid beta-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease.* Nat Med, 1996. **2**(8): p. 864-70.
96. Borchelt, D.R., et al., *Familial Alzheimer's disease-linked presenilin 1 variants elevate Abeta1-42/1-40 ratio in vitro and in vivo.* Neuron, 1996. **17**(5): p. 1005-13.
97. Duff, K., et al., *Increased amyloid-beta42(43) in brains of mice expressing mutant presenilin 1.* Nature, 1996. **383**(6602): p. 710-3.
98. Wakutani, Y., et al., *Novel amyloid precursor protein gene missense mutation (D678N) in probable familial Alzheimer's disease.* J Neurol Neurosurg Psychiatry, 2004. **75**(7): p. 1039-42.
99. Brouwers, N., K. Sleegers, and C. Van Broeckhoven, *Molecular genetics of Alzheimer's disease: an update.* Ann Med, 2008. **40**(8): p. 562-83.
100. Hendriks, L., et al., *Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the beta-amyloid precursor protein gene.* Nat Genet, 1992. **1**(3): p. 218-21.
101. Tagliavini F, R.G., Padovani A, Magoni M, Andora G, Sgarzi M, Bizzi A, Savoiardo M, Carella F, Morbin M, Giaccone G, Bugiani O., *A new βPP mutation related to hereditary cerebral haemorrhage.* Alzheimer's Reports, 1999. **2 Supp1**(S28).
102. Levy, E., et al., *Mutation of the Alzheimer's disease amyloid gene in hereditary cerebral hemorrhage, Dutch type.* Science, 1990. **248**(4959): p. 1124-6.
103. Van Broeckhoven, C., et al., *Amyloid beta protein precursor gene and hereditary cerebral hemorrhage with amyloidosis (Dutch).* Science, 1990. **248**(4959): p. 1120-2.

104. Fernandez-Madrid, I., et al., *Codon 618 variant of Alzheimer amyloid gene associated with inherited cerebral hemorrhage.* Ann Neurol, 1991. **30**(5): p. 730-3.
105. Kamino, K., et al., *Linkage and mutational analysis of familial Alzheimer disease kindreds for the APP gene region.* Am J Hum Genet, 1992. **51**(5): p. 998-1014.
106. Grabowski, T.J., et al., *Novel amyloid precursor protein mutation in an Iowa family with dementia and severe cerebral amyloid angiopathy.* Ann Neurol, 2001. **49**(6): p. 697-705.
107. Giaccone G, R.G., Morbin M, Tagliavini F, Bugiani O., *A713T mutation of the APP gene in an Italian family with Alzheimer disease and severe congophilic angiopathy.* . Neurobiology of Aging, 2002. **23 Supp 1**(320).
108. Rossi, G., et al., *A family with Alzheimer disease and strokes associated with A713T mutation of the APP gene.* Neurology, 2004. **63**(5): p. 910-2.
109. Armstrong, J., et al., *Familial Alzheimer disease associated with A713T mutation in APP.* Neurosci Lett, 2004. **370**(2-3): p. 241-3.
110. Bernardi, L., et al., *AbetaPP A713T mutation in late onset Alzheimer's disease with cerebrovascular lesions.* J Alzheimers Dis, 2009. **17**(2): p. 383-9.
111. Lindquist, S.G., et al., *Atypical early-onset Alzheimer's disease caused by the Iranian APP mutation.* J Neurol Sci, 2008. **268**(1-2): p. 124-30.
112. Pasalar, P., et al., *An Iranian family with Alzheimer's disease caused by a novel APP mutation (Thr714Ala).* Neurology, 2002. **58**(10): p. 1574-5.
113. Zekanowski, C., et al., *Mutations in presenilin 1, presenilin 2 and amyloid precursor protein genes in patients with early-onset Alzheimer's disease in Poland.* Exp Neurol, 2003. **184**(2): p. 991-6.
114. De Jonghe C, K.-S.S., Cruts M, Kleinert R, Vanderstichele H, Vanmechelen EJM, Kroisel P, Van Broeckhoven C., *Unusual Aβ amyloid deposition in Alzheimer's disease due tu an APP T714I mutation at the γ42-secretase site.* . Neurobiology of Aging 2000. **21**(S200).
115. Edwards-Lee, T., et al., *An African American family with early-onset Alzheimer disease and an APP (T714I) mutation.* Neurology, 2005. **64**(2): p. 377-9.
116. Ancolio, K., et al., *Unusual phenotypic alteration of beta amyloid precursor protein (betaAPP) maturation by a new Val-715 --> Met betaAPP-770 mutation responsible for probable early-onset Alzheimer's disease.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 4119-24.
117. Park, H.K., et al., *Identification of PSEN1 and APP gene mutations in Korean patients with early-onset Alzheimer's disease.* J Korean Med Sci, 2008. **23**(2): p. 213-7.
118. Cruts M, D.B., Kumar-Singh S, Rademakers R, Van den Broeck M, Stögbauer F, Van Broeckhoven, C., *Novel German APP V715A mutation associated with presenile Alzheimer's disease.* Neurobiology of Aging, 2002. **23**(S327).

119. Janssen JC, B.J., Campbell TA, Dickinson A, Fox NC, Harvey RJ, Houlden H, Rossor MN, Collinge J., *Early onset familial Alzheimer's disease: mutation frequency in 31 families.* Neurobiology of Aging, 2002. **23**(S311).
120. Eckman, C.B., et al., *A new pathogenic mutation in the APP gene (I716V) increases the relative proportion of A beta 42(43).* Hum Mol Genet, 1997. **6**(12): p. 2087-9.
121. Guerreiro, R.J., et al., *Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP.* Neurobiol Aging, 2010. **31**(5): p. 725-31.
122. Terreni L, F.S., Franceschi, Forloni G., *Novel pathogenic mutation in an Italian patient with familial Alzheimer's disease detected in APP gene.* Neurobiology of Aging, 2002. **23**(S319).
123. Campion, D., et al., *No founder effect in three novel Alzheimer's disease families with APP 717 Val-->Ile mutation. Clerget-darpoux. French Alzheimer's Disease Study Group.* J Med Genet, 1996. **33**(8): p. 661-4.
124. Brooks, W.S., et al., *A mutation in codon 717 of the amyloid precursor protein gene in an Australian family with Alzheimer's disease.* Neurosci Lett, 1995. **199**(3): p. 183-6.
125. Fidani, L., et al., *Screening for mutations in the open reading frame and promoter of the beta-amyloid precursor protein gene in familial Alzheimer's disease: identification of a further family with APP717 Val-->Ile.* Hum Mol Genet, 1992. **1**(3): p. 165-8.
126. Sorbi, S., et al., *APP717 and Alzheimer's disease in Italy.* Nat Genet, 1993. **4**(1): p. 10.
127. Yoshizawa, T., et al., *Screening of the mis-sense mutation producing the 717Val-->Ile substitution in the amyloid precursor protein in Japanese familial and sporadic Alzheimer's disease.* J Neurol Sci, 1993. **117**(1-2): p. 12-5.
128. Matsumura, Y., et al., *Japanese siblings with missense mutation (717Val --> Ile) in amyloid precursor protein of early-onset Alzheimer's disease.* Neurology, 1996. **46**(6): p. 1721-3.
129. Murrell, J.R., et al., *Early-onset Alzheimer disease caused by a new mutation (V717L) in the amyloid precursor protein gene.* Arch Neurol, 2000. **57**(6): p. 885-7.
130. Godbolt, A.K., et al., *A second family with familial AD and the V717L APP mutation has a later age at onset.* Neurology, 2006. **66**(4): p. 611-2.
131. Kwok JBJ, L.Q.-X., Hallupp M, Milward L, Whyte S, Schofield PR. , *Novel familial early-onset Alzheimer's disease mutation (Leu723Pro) in amyloid precursor protein (APP) gene increases production of 42(43) amino-acid isoform of amyloid beta peptide.* . Neurobiology of Aging, 1998. **19**(S91).
132. Theuns, J., et al., *Alzheimer dementia caused by a novel mutation located in the APP C-terminal intracytosolic fragment.* Hum Mutat, 2006. **27**(9): p. 888-96.
133. Gotz, J., et al., *Formation of neurofibrillary tangles in P301l tau transgenic mice induced by Abeta 42 fibrils.* Science, 2001. **293**(5534): p. 1491-5.

134. Hardy, J. and D. Allsop, *Amyloid deposition as the central event in the aetiology of Alzheimer's disease.* Trends Pharmacol Sci, 1991. **12**(10): p. 383-8.
135. Selkoe, D.J., *The molecular pathology of Alzheimer's disease.* Neuron, 1991. **6**(4): p. 487-98.
136. Hardy, J.A. and G.A. Higgins, *Alzheimer's disease: the amyloid cascade hypothesis.* Science, 1992. **256**(5054): p. 184-5.
137. Hardy, J. and D.J. Selkoe, *The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics.* Science, 2002. **297**(5580): p. 353-6.
138. Vetrivel, K.S. and G. Thinakaran, *Amyloidogenic processing of beta-amyloid precursor protein in intracellular compartments.* Neurology, 2006. **66**(2 Suppl 1): p. S69-73.
139. Anandatheerthavarada, H.K., et al., *Mitochondrial targeting and a novel transmembrane arrest of Alzheimer's amyloid precursor protein impairs mitochondrial function in neuronal cells.* J Cell Biol, 2003. **161**(1): p. 41-54.
140. Vetrivel, K.S., et al., *Association of gamma-secretase with lipid rafts in post-Golgi and endosome membranes.* J Biol Chem, 2004. **279**(43): p. 44945-54.
141. Golde, T.E., et al., *Processing of the amyloid protein precursor to potentially amyloidogenic derivatives.* Science, 1992. **255**(5045): p. 728-30.
142. Kamenetz, F., et al., *APP processing and synaptic function.* Neuron, 2003. **37**(6): p. 925-37.
143. Cirrito, J.R., et al., *Endocytosis is required for synaptic activity-dependent release of amyloid-beta in vivo.* Neuron, 2008. **58**(1): p. 42-51.
144. Cirrito, J.R., et al., *Synaptic activity regulates interstitial fluid amyloid-beta levels in vivo.* Neuron, 2005. **48**(6): p. 913-22.
145. Taylor, C.J., et al., *Endogenous secreted amyloid precursor protein-alpha regulates hippocampal NMDA receptor function, long-term potentiation and spatial memory.* Neurobiol Dis, 2008. **31**(2): p. 250-60.
146. Hardy, J., *Toward Alzheimer therapies based on genetic knowledge.* Annu Rev Med, 2004. **55**: p. 15-25.
147. Chow, V.W., et al., *An overview of APP processing enzymes and products.* Neuromolecular Med, 2010. **12**(1): p. 1-12.
148. Ishida, A., et al., *Secreted form of beta-amyloid precursor protein shifts the frequency dependency for induction of LTD, and enhances LTP in hippocampal slices.* Neuroreport, 1997. **8**(9-10): p. 2133-7.
149. Gakhar-Koppole, N., et al., *Activity requires soluble amyloid precursor protein alpha to promote neurite outgrowth in neural stem cell-derived neurons via activation of the MAPK pathway.* Eur J Neurosci, 2008. **28**(5): p. 871-82.
150. Barger, S.W. and A.D. Harmon, *Microglial activation by Alzheimer amyloid precursor protein and modulation by apolipoprotein E.* Nature, 1997. **388**(6645): p. 878-81.
151. Barger, S.W. and A.S. Basile, *Activation of microglia by secreted amyloid precursor protein evokes release of glutamate by cystine exchange and attenuates synaptic function.* J Neurochem, 2001. **76**(3): p. 846-54.

152. Li, H., et al., *Soluble amyloid precursor protein (APP) regulates transthyretin and Klotho gene expression without rescuing the essential function of APP.* Proc Natl Acad Sci U S A, 2010. **107**(40): p. 17362-7.
153. Nabeshima, Y., *Klotho: a fundamental regulator of aging.* Ageing Res Rev, 2002. **1**(4): p. 627-38.
154. Du, J. and R.M. Murphy, *Characterization of the interaction of beta-amyloid with transthyretin monomers and tetramers.* Biochemistry, 2010. **49**(38): p. 8276-89.
155. Haass, C., et al., *beta-Amyloid peptide and a 3-kDa fragment are derived by distinct cellular mechanisms.* J Biol Chem, 1993. **268**(5): p. 3021-4.
156. Lalowski, M., et al., *The "nonamyloidogenic" p3 fragment (amyloid beta17-42) is a major constituent of Down's syndrome cerebellar preamyloid.* J Biol Chem, 1996. **271**(52): p. 33623-31.
157. Wei, W., et al., *Abeta 17-42 in Alzheimer's disease activates JNK and caspase-8 leading to neuronal apoptosis.* Brain, 2002. **125**(Pt 9): p. 2036-43.
158. Slomnicki, L.P. and W. Lesniak, *A putative role of the Amyloid Precursor Protein Intracellular Domain (AICD) in transcription.* Acta Neurobiol Exp (Wars), 2008. **68**(2): p. 219-28.
159. Goodger, Z.V., et al., *Nuclear signaling by the APP intracellular domain occurs predominantly through the amyloidogenic processing pathway.* J Cell Sci, 2009. **122**(Pt 20): p. 3703-14.
160. Koike, H., et al., *Membrane-anchored metalloprotease MDC9 has an alpha-secretase activity responsible for processing the amyloid precursor protein.* Biochem J, 1999. **343 Pt 2**: p. 371-5.
161. Lammich, S., et al., *Constitutive and regulated alpha-secretase cleavage of Alzheimer's amyloid precursor protein by a disintegrin metalloprotease.* Proc Natl Acad Sci U S A, 1999. **96**(7): p. 3922-7.
162. Buxbaum, J.D., et al., *Evidence that tumor necrosis factor alpha converting enzyme is involved in regulated alpha-secretase cleavage of the Alzheimer amyloid protein precursor.* J Biol Chem, 1998. **273**(43): p. 27765-7.
163. Tanabe, C., et al., *ADAM19 is tightly associated with constitutive Alzheimer's disease APP alpha-secretase in A172 cells.* Biochem Biophys Res Commun, 2007. **352**(1): p. 111-7.
164. Kim, M., et al., *Potential late-onset Alzheimer's disease-associated mutations in the ADAM10 gene attenuate {alpha}-secretase activity.* Hum Mol Genet, 2009. **18**(20): p. 3987-96.
165. Prinzen, C., et al., *Genomic structure and functional characterization of the human ADAM10 promoter.* FASEB J, 2005. **19**(11): p. 1522-4.
166. Harold, D., et al., *Interaction between the ADAM12 and SH3MD1 genes may confer susceptibility to late-onset Alzheimer's disease.* Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(4): p. 448-52.
167. Laumet, G., et al., *A study of the association between the ADAM12 and SH3PXD2A (SH3MD1) genes and Alzheimer's disease.* Neurosci Lett, 2010. **468**(1): p. 1-2.

168. Cong, L. and J. Jia, *Promoter polymorphisms which regulate ADAM9 transcription are protective against sporadic Alzheimer's disease.* Neurobiol Aging, 2009.
169. De Strooper, B. and W. Annaert, *Proteolytic processing and cell biological functions of the amyloid precursor protein.* J Cell Sci, 2000. **113 (Pt 11)**: p. 1857-70.
170. De Strooper, B., R. Vassar, and T. Golde, *The secretases: enzymes with therapeutic potential in Alzheimer disease.* Nat Rev Neurol, 2010. **6**(2): p. 99-107.
171. Selkoe, D.J., et al., *The role of APP processing and trafficking pathways in the formation of amyloid beta-protein.* Ann N Y Acad Sci, 1996. **777**: p. 57-64.
172. Thinakaran, G., et al., *Metabolism of the "Swedish" amyloid precursor protein variant in neuro2a (N2a) cells. Evidence that cleavage at the "beta-secretase" site occurs in the golgi apparatus.* J Biol Chem, 1996. **271**(16): p. 9390-7.
173. Cai, H., et al., *BACE1 is the major beta-secretase for generation of Abeta peptides by neurons.* Nat Neurosci, 2001. **4**(3): p. 233-4.
174. Lee, S.F., et al., *Mammalian APH-1 interacts with presenilin and nicastrin and is required for intramembrane proteolysis of amyloid-beta precursor protein and Notch.* J Biol Chem, 2002. **277**(47): p. 45013-9.
175. Steiner, H., et al., *PEN-2 is an integral component of the gamma-secretase complex required for coordinated expression of presenilin and nicastrin.* J Biol Chem, 2002. **277**(42): p. 39062-5.
176. Greenfield, J.P., et al., *Endoplasmic reticulum and trans-Golgi network generate distinct populations of Alzheimer beta-amyloid peptides.* Proc Natl Acad Sci U S A, 1999. **96**(2): p. 742-7.
177. Chyung, J.H., D.M. Raper, and D.J. Selkoe, *Gamma-secretase exists on the plasma membrane as an intact complex that accepts substrates and effects intramembrane cleavage.* J Biol Chem, 2005. **280**(6): p. 4383-92.
178. De Strooper, B., et al., *A presenilin-1-dependent gamma-secretase-like protease mediates release of Notch intracellular domain.* Nature, 1999. **398**(6727): p. 518-22.
179. Wong, P.C., et al., *Presenilin 1 is required for Notch1 and DII1 expression in the paraxial mesoderm.* Nature, 1997. **387**(6630): p. 288-92.
180. Wong, G.T., et al., *Chronic treatment with the gamma-secretase inhibitor LY-411,575 inhibits beta-amyloid peptide production and alters lymphopoiesis and intestinal cell differentiation.* J Biol Chem, 2004. **279**(13): p. 12876-82.
181. Searfoss, G.H., et al., *Adipsin, a biomarker of gastrointestinal toxicity mediated by a functional gamma-secretase inhibitor.* J Biol Chem, 2003. **278**(46): p. 46107-16.
182. Huppert, S.S., et al., *Embryonic lethality in mice homozygous for a processing-deficient allele of Notch1.* Nature, 2000. **405**(6789): p. 966-70.
183. Wang, B., et al., *Gamma-secretase gene mutations in familial acne inversa.* Science, 2010. **330**(6007): p. 1065.

184. He, G., et al., *Gamma-secretase activating protein is a therapeutic target for Alzheimer's disease.* Nature, 2010. **467**(7311): p. 95-8.
185. Serneels, L., et al., *gamma-Secretase heterogeneity in the Aph1 subunit: relevance for Alzheimer's disease.* Science, 2009. **324**(5927): p. 639-42.
186. Deane, R., et al., *Clearance of amyloid-beta peptide across the blood-brain barrier: implication for therapies in Alzheimer's disease.* CNS Neurol Disord Drug Targets, 2009. **8**(1): p. 16-30.
187. Citron, M., *Alzheimer's disease: strategies for disease modification.* Nat Rev Drug Discov, 2010. **9**(5): p. 387-98.
188. Skogen, B. and J.B. Natvig, *Degradation of amyloid proteins by different serine proteases.* Scand J Immunol, 1981. **14**(4): p. 389-96.
189. Backstrom, J.R., et al., *Matrix metalloproteinase-9 (MMP-9) is synthesized in neurons of the human hippocampus and is capable of degrading the amyloid-beta peptide (1-40).* J Neurosci, 1996. **16**(24): p. 7910-9.
190. Kurochkin, I.V. and S. Goto, *Alzheimer's beta-amyloid peptide specifically interacts with and is degraded by insulin degrading enzyme.* FEBS Lett, 1994. **345**(1): p. 33-7.
191. Eckman, E.A., D.K. Reed, and C.B. Eckman, *Degradation of the Alzheimer's amyloid beta peptide by endothelin-converting enzyme.* J Biol Chem, 2001. **276**(27): p. 24540-8.
192. Hu, J., et al., *Angiotensin-converting enzyme degrades Alzheimer amyloid beta-peptide (A beta); retards A beta aggregation, deposition, fibril formation; and inhibits cytotoxicity.* J Biol Chem, 2001. **276**(51): p. 47863-8.
193. Howell, S., J. Nalbantoglu, and P. Crine, *Neutral endopeptidase can hydrolyze beta-amyloid(1-40) but shows no effect on beta-amyloid precursor protein metabolism.* Peptides, 1995. **16**(4): p. 647-52.
194. Fiala, M., et al., *Phagocytosis of amyloid-beta and inflammation: two faces of innate immunity in Alzheimer's disease.* J Alzheimers Dis, 2007. **11**(4): p. 457-63.
195. Fiala, M., et al., *Ineffective phagocytosis of amyloid-beta by macrophages of Alzheimer's disease patients.* J Alzheimers Dis, 2005. **7**(3): p. 221-32; discussion 255-62.
196. Wilcock, D.M., et al., *Intracranially administered anti-Abeta antibodies reduce beta-amyloid deposition by mechanisms both independent of and associated with microglial activation.* J Neurosci, 2003. **23**(9): p. 3745-51.
197. Schenk, D., et al., *Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse.* Nature, 1999. **400**(6740): p. 173-7.
198. Janus, C., et al., *A beta peptide immunization reduces behavioural impairment and plaques in a model of Alzheimer's disease.* Nature, 2000. **408**(6815): p. 979-82.
199. Hock, C., et al., *Generation of antibodies specific for beta-amyloid by vaccination of patients with Alzheimer disease.* Nat Med, 2002. **8**(11): p. 1270-5.
200. Hock, C., et al., *Antibodies against beta-amyloid slow cognitive decline in Alzheimer's disease.* Neuron, 2003. **38**(4): p. 547-54.

201. Nitsch, R.M. and C. Hock, *Targeting beta-amyloid pathology in Alzheimer's disease with Abeta immunotherapy.* Neurotherapeutics, 2008. **5**(3): p. 415-20.
202. Vellas, B., et al., *Long-term follow-up of patients immunized with AN1792: reduced functional decline in antibody responders.* Curr Alzheimer Res, 2009. **6**(2): p. 144-51.
203. Graff-Radford, N.R., et al., *Association of low plasma Abeta42/Abeta40 ratios with increased imminent risk for mild cognitive impairment and Alzheimer disease.* Arch Neurol, 2007. **64**(3): p. 354-62.
204. Ertekin-Taner, N., et al., *Heritability of plasma amyloid beta in typical late-onset Alzheimer's disease pedigrees.* Genet Epidemiol, 2001. **21**(1): p. 19-30.
205. Ertekin-Taner, N., et al., *Genetic variants in a haplotype block spanning IDE are significantly associated with plasma Abeta42 levels and risk for Alzheimer disease.* Hum Mutat, 2004. **23**(4): p. 334-42.
206. Ertekin-Taner, N., et al., *Fine mapping of the alpha-T catenin gene to a quantitative trait locus on chromosome 10 in late-onset Alzheimer's disease pedigrees.* Hum Mol Genet, 2003. **12**(23): p. 3133-43.
207. Ertekin-Taner, N., et al., *Elevated amyloid beta protein (Abeta42) and late onset Alzheimer's disease are associated with single nucleotide polymorphisms in the urokinase-type plasminogen activator gene.* Hum Mol Genet, 2005. **14**(3): p. 447-60.
208. Gatz, M., et al., *Role of genes and environments for explaining Alzheimer disease.* Arch Gen Psychiatry, 2006. **63**(2): p. 168-74.
209. Nowotny, P., et al., *Association studies testing for risk for late-onset Alzheimer's disease with common variants in the beta-amyloid precursor protein (APP).* Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(4): p. 469-74.
210. Guyant-Marechal, L., et al., *Variations in the APP gene promoter region and risk of Alzheimer disease.* Neurology, 2007. **68**(9): p. 684-7.
211. Lv, H., L. Jia, and J. Jia, *Promoter polymorphisms which modulate APP expression may increase susceptibility to Alzheimer's disease.* Neurobiol Aging, 2008. **29**(2): p. 194-202.
212. Lambert, J.C., et al., *The -48 C/T polymorphism in the presenilin 1 promoter is associated with an increased risk of developing Alzheimer's disease and an increased Abeta load in brain.* J Med Genet, 2001. **38**(6): p. 353-5.
213. Liu, Z. and J. Jia, *The association of the regulatory region of the presenilin-2 gene with Alzheimer's disease in the Northern Han Chinese population.* J Neurol Sci, 2008. **264**(1-2): p. 38-42.
214. Matsubara-Tsutsui, M., et al., *The 4,752 C/T polymorphism in the presenilin 1 gene increases the risk of Alzheimer's disease in apolipoprotein E4 carriers.* Intern Med, 2002. **41**(10): p. 823-8.
215. Martinez-Garcia, A., et al., *Presenilin 1 polymorphism associated with Alzheimer's disease in apolipoprotein E4 carriers.* Dement Geriatr Cogn Disord, 2008. **26**(5): p. 440-4.

216. Pericak-Vance, M.A., et al., *Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage.* Am J Hum Genet, 1991. **48**(6): p. 1034-50.
217. Corder, E.H., et al., *Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families.* Science, 1993. **261**(5123): p. 921-3.
218. Saunders, A.M., et al., *Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease.* Neurology, 1993. **43**(8): p. 1467-72.
219. Strittmatter, W.J., et al., *Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease.* Proc Natl Acad Sci U S A, 1993. **90**(5): p. 1977-81.
220. Schmechel, D.E., et al., *Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease.* Proc Natl Acad Sci U S A, 1993. **90**(20): p. 9649-53.
221. Strittmatter, W.J., et al., *Binding of human apolipoprotein E to synthetic amyloid beta peptide: isoform-specific effects and implications for late-onset Alzheimer disease.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 8098-102.
222. Ioannidis, J.P., et al., *Assessment of cumulative evidence on genetic associations: interim guidelines.* Int J Epidemiol, 2008. **37**(1): p. 120-32.
223. Khoury, M.J., et al., *Genome-wide association studies, field synopses, and the development of the knowledge base on genetic variation and human diseases.* Am J Epidemiol, 2009. **170**(3): p. 269-79.
224. Goring, H.H., J.D. Terwilliger, and J. Blangero, *Large upward bias in estimation of locus-specific effects from genomewide scans.* Am J Hum Genet, 2001. **69**(6): p. 1357-69.
225. Panagiotou, O.A., E. Evangelou, and J.P. Ioannidis, *Genome-wide significant associations for variants with minor allele frequency of 5% or less--an overview: A HuGE review.* Am J Epidemiol, 2010. **172**(8): p. 869-89.
226. Koch, W., et al., *TaqMan systems for genotyping of disease-related polymorphisms present in the gene encoding apolipoprotein E.* Clin Chem Lab Med, 2002. **40**(11): p. 1123-31.
227. Corder, E.H., et al., *Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease.* Nat Genet, 1994. **7**(2): p. 180-4.
228. Seshadri, S., et al., *Genome-wide analysis of genetic loci associated with Alzheimer disease.* JAMA, 2010. **303**(18): p. 1832-40.
229. Puglielli, L., R.E. Tanzi, and D.M. Kovacs, *Alzheimer's disease: the cholesterol connection.* Nat Neurosci, 2003. **6**(4): p. 345-51.
230. Wollmer, M.A., *Cholesterol-related genes in Alzheimer's disease.* Biochim Biophys Acta, 2010. **1801**(8): p. 762-73.
231. Vance, J.E. and H. Hayashi, *Formation and function of apolipoprotein E-containing lipoproteins in the nervous system.* Biochim Biophys Acta, 2010. **1801**(8): p. 806-18.
232. Reiman, E.M., et al., *GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers.* Neuron, 2007. **54**(5): p. 713-20.

233. Chapuis, J., et al., *Association study of the GAB2 gene with the risk of developing Alzheimer's disease.* Neurobiol Dis, 2008. **30**(1): p. 103-6.
234. Sleegers, K., et al., *Common variation in GRB-associated Binding Protein 2 (GAB2) and increased risk for Alzheimer dementia.* Hum Mutat, 2009. **30**(2): p. E338-44.
235. Grupe, A., et al., *Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants.* Hum Mol Genet, 2007. **16**(8): p. 865-73.
236. Coon, K.D., et al., *A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease.* J Clin Psychiatry, 2007. **68**(4): p. 613-8.
237. Li, H., et al., *Candidate single-nucleotide polymorphisms from a genomewide association study of Alzheimer disease.* Arch Neurol, 2008. **65**(1): p. 45-53.
238. Abraham, R., et al., *A genome-wide association study for late-onset Alzheimer's disease using DNA pooling.* BMC Med Genomics, 2008. **1**: p. 44.
239. Bertram, L., et al., *Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE.* Am J Hum Genet, 2008. **83**(5): p. 623-32.
240. Beecham, G.W., et al., *Genome-wide association study implicates a chromosome 12 risk locus for late-onset Alzheimer disease.* Am J Hum Genet, 2009. **84**(1): p. 35-43.
241. Carrasquillo, M.M., et al., *Genetic variation in PCDH11X is associated with susceptibility to late-onset Alzheimer's disease.* Nat Genet, 2009. **41**(2): p. 192-8.
242. Lambert, J.C., et al., *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1094-9.
243. Harold, D., et al., *Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1088-93.
244. Naj, A.C., et al., *Dementia revealed: novel chromosome 6 locus for late-onset Alzheimer disease provides genetic evidence for folate-pathway abnormalities.* PLoS Genet, 2010. **6**(9).
245. Lee, J.H., et al., *Identification of Novel Loci for Alzheimer Disease and Replication of CLU, PICALM, and BIN1 in Caribbean Hispanic Individuals.* Arch Neurol, 2010.
246. Bertram, L., C.M. Lill, and R.E. Tanzi, *The genetics of Alzheimer disease: back to the future.* Neuron, 2010. **68**(2): p. 270-81.
247. Corneveaux, J.J., et al., *Association of CR1, CLU and PICALM with Alzheimer's disease in a cohort of clinically characterized and neuropathologically verified individuals.* Hum Mol Genet, 2010. **19**(16): p. 3295-301.
248. Yu, J.T., et al., *Implication of CLU gene polymorphisms in Chinese patients with Alzheimer's disease.* Clin Chim Acta, 2010. **411**(19-20): p. 1516-9.

249. Jun, G., et al., *Meta-analysis Confirms CR1, CLU, and PICALM as Alzheimer Disease Risk Loci and Reveals Interactions With APOE Genotypes.* Arch Neurol, 2010.
250. Carrasquillo, M.M., et al., *Replication of CLU, CR1, and PICALM associations with alzheimer disease.* Arch Neurol, 2010. **67**(8): p. 961-4.
251. Rogaeva, E., et al., *The neuronal sortilin-related receptor SORL1 is genetically associated with Alzheimer disease.* Nat Genet, 2007. **39**(2): p. 168-77.
252. Infante, J., et al., *Gene-gene interaction between interleukin-1A and interleukin-8 increases Alzheimer's disease risk.* J Neurol, 2004. **251**(4): p. 482-3.
253. Thorisson, G.A., et al., *The International HapMap Project Web site.* Genome Res, 2005. **15**(11): p. 1592-3.
254. Durbin, R.M., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.
255. Guo, X., et al., *Genome-wide linkage of plasma adiponectin reveals a major locus on chromosome 3q distinct from the adiponectin structural gene: the IRAS family study.* Diabetes, 2006. **55**(6): p. 1723-30.
256. Pearlson, G.D. and V.D. Calhoun, *Convergent approaches for defining functional imaging endophenotypes in schizophrenia.* Front Hum Neurosci, 2009. **3**: p. 37.
257. Lien, Y.J., et al., *A genome-wide quantitative trait loci scan of neurocognitive performances in families with schizophrenia.* Genes Brain Behav, 2010. **9**(7): p. 695-702.
258. Wright, A.F., A.D. Carothers, and M. Pirastu, *Population choice in mapping genes for complex diseases.* Nat Genet, 1999. **23**(4): p. 397-404.
259. Kwitek-Black, A.E., et al., *Linkage of Bardet-Biedl syndrome to chromosome 16q and evidence for non-allelic genetic heterogeneity.* Nat Genet, 1993. **5**(4): p. 392-6.
260. Ekelund, J., et al., *Chromosome 1 loci in Finnish schizophrenia families.* Hum Mol Genet, 2001. **10**(15): p. 1611-7.
261. Helgason, A., et al., *Refining the impact of TCF7L2 gene variants on type 2 diabetes and adaptive evolution.* Nat Genet, 2007. **39**(2): p. 218-25.
262. Herzberg, I., et al., *Convergent linkage evidence from two Latin-American population isolates supports the presence of a susceptibility locus for bipolar disorder in 5q31-34.* Hum Mol Genet, 2006. **15**(21): p. 3146-53.
263. Hovatta, I., et al., *A genomewide screen for schizophrenia genes in an isolated Finnish subpopulation, suggesting multiple susceptibility loci.* Am J Hum Genet, 1999. **65**(4): p. 1114-24.
264. Vitart, V., et al., *3000 years of solitude: extreme differentiation in the island isolates of Dalmatia, Croatia.* Eur J Hum Genet, 2006. **14**(4): p. 478-87.
265. Rudan, I., H. Campbell, and P. Rudan, *Genetic epidemiological studies of eastern Adriatic Island isolates, Croatia: objective and strategies.* Coll Antropol, 1999. **23**(2): p. 531-46.
266. Rudan, I., et al., *"10001 Dalmatians:" Croatia launches its national biobank.* Croat Med J, 2009. **50**(1): p. 4-6.

267. Barac, L., et al., *Y chromosomal heritage of Croatian population and its island isolates.* Eur J Hum Genet, 2003. **11**(7): p. 535-42.
268. Tolk, H.V., et al., *MtDNA haplogroups in the populations of Croatian Adriatic Islands.* Coll Antropol, 2000. **24**(2): p. 267-80.
269. Navarro, P., et al., *Genetic comparison of a Croatian isolate and CEPH European founders.* Genet Epidemiol, 2010. **34**(2): p. 140-5.
270. Polasek, O., et al., *Genome-wide association study of anthropometric traits in Korcula Island, Croatia.* Croat Med J, 2009. **50**(1): p. 7-16.
271. Zemunik, T., et al., *Genome-wide association study of biochemical traits in Korcula Island, Croatia.* Croat Med J, 2009. **50**(1): p. 23-33.
272. Vitart, V., et al., *SLC2A9 is a newly identified urate transporter influencing serum urate concentration, urate excretion and gout.* Nat Genet, 2008. **40**(4): p. 437-42.
273. Doring, A., et al., *SLC2A9 influences uric acid concentrations with pronounced sex-specific effects.* Nat Genet, 2008. **40**(4): p. 430-6.
274. Murphy, M.P., *How mitochondria produce reactive oxygen species.* Biochem J, 2009. **417**(1): p. 1-13.
275. Harman, D., *The biologic clock: the mitochondria?* J Am Geriatr Soc, 1972. **20**(4): p. 145-7.
276. Miquel, J., et al., *Mitochondrial role in cell aging.* Exp Gerontol, 1980. **15**(6): p. 575-91.
277. Wright, A.F., M.P. Murphy, and D.M. Turnbull, *Do organellar genomes function as long-term redox damage sensors?* Trends Genet, 2009. **25**(6): p. 253-61.
278. Wallace, D.C., *Mitochondrial DNA mutations in disease and aging.* Environ Mol Mutagen, 2010. **51**(5): p. 440-50.
279. Wang, J., et al., *Increased oxidative damage in nuclear and mitochondrial DNA in Alzheimer's disease.* J Neurochem, 2005. **93**(4): p. 953-62.
280. Parsons, T.J., et al., *A high observed substitution rate in the human mitochondrial DNA control region.* Nat Genet, 1997. **15**(4): p. 363-8.
281. Loeb, L.A., D.C. Wallace, and G.M. Martin, *The mitochondrial theory of aging and its relationship to reactive oxygen species damage and somatic mtDNA mutations.* Proc Natl Acad Sci U S A, 2005. **102**(52): p. 18769-70.
282. Trifunovic, A., et al., *Premature ageing in mice expressing defective mitochondrial DNA polymerase.* Nature, 2004. **429**(6990): p. 417-23.
283. Kujoth, G.C., et al., *Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging.* Science, 2005. **309**(5733): p. 481-4.
284. Schriner, S.E., et al., *Extension of murine life span by overexpression of catalase targeted to mitochondria.* Science, 2005. **308**(5730): p. 1909-11.
285. de Moura, M.B., L.S. dos Santos, and B. Van Houten, *Mitochondrial dysfunction in neurodegenerative diseases and cancer.* Environ Mol Mutagen, 2010. **51**(5): p. 391-405.
286. Smigrodzki, R.M. and S.M. Khan, *Mitochondrial microheteroplasmy and a theory of aging and age-related disease.* Rejuvenation Res, 2005. **8**(3): p. 172-98.

287. Coskun, P.E., M.F. Beal, and D.C. Wallace, *Alzheimer's brains harbor somatic mtDNA control-region mutations that suppress mitochondrial transcription and replication.* Proc Natl Acad Sci U S A, 2004. **101**(29): p. 10726-31.
288. Kraytsberg, Y., et al., *Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons.* Nat Genet, 2006. **38**(5): p. 518-20.
289. Lakatos, A., et al., *Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort.* Neurobiol Aging, 2010. **31**(8): p. 1355-63.
290. Mecocci, P., U. MacGarvey, and M.F. Beal, *Oxidative damage to mitochondrial DNA is increased in Alzheimer's disease.* Ann Neurol, 1994. **36**(5): p. 747-51.
291. Tanaka, N., et al., *Mitochondrial DNA variants in a Japanese population of patients with Alzheimer's disease.* Mitochondrion, 2010. **10**(1): p. 32-7.
292. Corral-Debrinski, M., et al., *Mitochondrial DNA deletions in human brain: regional variability and increase with advanced age.* Nat Genet, 1992. **2**(4): p. 324-9.
293. Mawrin, C., et al., *Region-specific analysis of mitochondrial DNA deletions in neurodegenerative disorders in humans.* Neurosci Lett, 2004. **357**(2): p. 111-4.
294. Zhang, J., et al., *The mitochondrial common deletion in Parkinson's disease and related movement disorders.* Parkinsonism Relat Disord, 2002. **8**(3): p. 165-70.
295. Lin, M.T., et al., *High aggregate burden of somatic mtDNA point mutations in aging and Alzheimer's disease brain.* Hum Mol Genet, 2002. **11**(2): p. 133-45.
296. Smigrodzki, R., J. Parks, and W.D. Parker, *High frequency of mitochondrial complex I mutations in Parkinson's disease and aging.* Neurobiol Aging, 2004. **25**(10): p. 1273-81.
297. Chinnery, P.F., et al., *Point mutations of the mtDNA control region in normal and neurodegenerative human brains.* Am J Hum Genet, 2001. **68**(2): p. 529-32.
298. Elson, J.L., et al., *Does the mitochondrial genome play a role in the etiology of Alzheimer's disease?* Hum Genet, 2006. **119**(3): p. 241-54.
299. Blass, J.P., *Brain metabolism and brain disease: is metabolic deficiency the proximate cause of Alzheimer dementia?* J Neurosci Res, 2001. **66**(5): p. 851-6.
300. Mosconi, L., et al., *Declining brain glucose metabolism in normal individuals with a maternal history of Alzheimer disease.* Neurology, 2009. **72**(6): p. 513-20.
301. Mosconi, L., et al., *FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease.* Eur J Nucl Med Mol Imaging, 2009. **36**(5): p. 811-22.

302. Mutisya, E.M., A.C. Bowling, and M.F. Beal, *Cortical cytochrome oxidase activity is reduced in Alzheimer's disease.* J Neurochem, 1994. **63**(6): p. 2179-84.
303. Maurer, I., S. Zierz, and H.J. Moller, *A selective defect of cytochrome c oxidase is present in brain of Alzheimer disease patients.* Neurobiol Aging, 2000. **21**(3): p. 455-62.
304. Cardoso, S.M., et al., *Mitochondria dysfunction of Alzheimer's disease cybrids enhances Abeta toxicity.* J Neurochem, 2004. **89**(6): p. 1417-26.
305. Devi, L., et al., *Accumulation of amyloid precursor protein in the mitochondrial import channels of human Alzheimer's disease brain is associated with mitochondrial dysfunction.* J Neurosci, 2006. **26**(35): p. 9057-68.
306. Lustbader, J.W., et al., *ABAD directly links Abeta to mitochondrial toxicity in Alzheimer's disease.* Science, 2004. **304**(5669): p. 448-52.
307. Launer, L.J., et al., *Rates and risk factors for dementia and Alzheimer's disease: results from EURODEM pooled analyses. EURODEM Incidence Research Group and Work Groups. European Studies of Dementia.* Neurology, 1999. **52**(1): p. 78-84.
308. Kamphuis, P.J. and P. Scheltens, *Can nutrients prevent or delay onset of Alzheimer's disease?* J Alzheimers Dis, 2010. **20**(3): p. 765-75.
309. Sofi, F., et al., *Effectiveness of the Mediterranean diet: can it help delay or prevent Alzheimer's disease?* J Alzheimers Dis, 2010. **20**(3): p. 795-801.
310. Mortimer, J.A., et al., *Head trauma as a risk factor for Alzheimer's disease: a collaborative re-analysis of case-control studies. EURODEM Risk Factors Research Group.* Int J Epidemiol, 1991. **20 Suppl 2**: p. S28-35.
311. Van Duijn, C.M., et al., *Interaction between genetic and environmental risk factors for Alzheimer's disease: a reanalysis of case-control studies. EURODEM Risk Factors Research Group.* Genet Epidemiol, 1994. **11**(6): p. 539-51.
312. Weuve J, M.M., Blacker D. *The AlzRisk Database. Alzheimer research forum*. [cited 2010 12-20-2010]; Available from: www.alzforum.org.
313. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
314. Kanehisa, M., et al., *KEGG for representation and analysis of molecular networks involving diseases and drugs.* Nucleic Acids Res, 2010. **38**(Database issue): p. D355-60.
315. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG.* Nucleic Acids Res, 2006. **34**(Database issue): p. D354-7.
316. Dupuis, J., et al., *New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk.* Nat Genet, 2010. **42**(2): p. 105-16.
317. Smith, N.L., et al., *Novel associations of multiple genetic loci with plasma levels of factor VII, factor VIII, and von Willebrand factor: The CHARGE (Cohorts for Heart and Aging Research in Genome Epidemiology) Consortium.* Circulation, 2010. **121**(12): p. 1382-92.

318. Aulchenko, Y.S., et al., *GenABEL: an R library for genome-wide association analysis.* Bioinformatics, 2007. **23**(10): p. 1294-6.
319. Shmulewitz, D. and S.C. Heath, *Genome scans for Q1 and Q2 on general population replicates using Loki.* Genet Epidemiol, 2001. **21 Suppl 1**: p. S686-91.
320. Yuan, B., et al., *Linkage of a gene for familial hypobetalipoproteinemia to chromosome 3p21.1-22.* Am J Hum Genet, 2000. **66**(5): p. 1699-704.
321. Almasy, L. and J. Blangero, *Multipoint quantitative-trait linkage analysis in general pedigrees.* Am J Hum Genet, 1998. **62**(5): p. 1198-211.
322. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.
323. Figgins, J.A., et al., *Association studies of 22 candidate SNPs with late-onset Alzheimer's disease.* Am J Med Genet B Neuropsychiatr Genet, 2009. **150B**(4): p. 520-6.
324. Schjeide, B.M., et al., *Assessment of Alzheimer's disease case-control associations using family-based methods.* Neurogenetics, 2009. **10**(1): p. 19-25.
325. Law, S.F., et al., *Human enhancer of filamentation 1, a novel p130cas-like docking protein, associates with focal adhesion kinase and induces pseudohyphal growth in Saccharomyces cerevisiae.* Mol Cell Biol, 1996. **16**(7): p. 3327-37.
326. Minegishi, M., et al., *Structure and function of Cas-L, a 105-kD Crk-associated substrate-related protein that is involved in beta 1 integrin-mediated signaling in lymphocytes.* J Exp Med, 1996. **184**(4): p. 1365-75.
327. Sanz-Moreno, V., et al., *Rac activation and inactivation control plasticity of tumor cell movement.* Cell, 2008. **135**(3): p. 510-23.
328. Kim, M., et al., *Comparative oncogenomics identifies NEDD9 as a melanoma metastasis gene.* Cell, 2006. **125**(7): p. 1269-81.
329. Xia, D., et al., *HEF1 is a crucial mediator of the proliferative effects of prostaglandin E(2) on colon cancer cells.* Cancer Res, 2010. **70**(2): p. 824-31.
330. Chapuis, J., et al., *Association study of the NEDD9 gene with the risk of developing Alzheimer's and Parkinson's disease.* Hum Mol Genet, 2008. **17**(18): p. 2863-7.
331. Li, Y., et al., *Evidence that common variation in NEDD9 is associated with susceptibility to late-onset Alzheimer's and Parkinson's disease.* Hum Mol Genet, 2008. **17**(5): p. 759-67.
332. Tedde, A., et al., *Different implication of NEDD9 genetic variant in early and late-onset Alzheimer's disease.* Neurosci Lett, 2010. **477**(3): p. 121-3.
333. Ashley-Koch, A.E., et al., *An autosomal genomic screen for dementia in an extended Amish family.* Neurosci Lett, 2005. **379**(3): p. 199-204.
334. Blacker, D., et al., *Results of a high-resolution genome screen of 437 Alzheimer's disease families.* Hum Mol Genet, 2003. **12**(1): p. 23-32.
335. Curtis, D., B.V. North, and P.C. Sham, *A novel method of two-locus linkage analysis applied to a genome scan for late onset Alzheimer's disease.* Ann Hum Genet, 2001. **65**(Pt 5): p. 473-81.

336. Giedraitis, V., et al., *New Alzheimer's disease locus on chromosome 8.* J Med Genet, 2006. **43**(12): p. 931-5.
337. Hahs, D.W., et al., *A genome-wide linkage analysis of dementia in the Amish.* Am J Med Genet B Neuropsychiatr Genet, 2006. **141B**(2): p. 160-6.
338. Hamshere, M.L., et al., *Genome-wide linkage analysis of 723 affected relative pairs with late-onset Alzheimer's disease.* Hum Mol Genet, 2007. **16**(22): p. 2703-12.
339. Kehoe, P., et al., *A full genome scan for late onset Alzheimer's disease.* Hum Mol Genet, 1999. **8**(2): p. 237-45.
340. Lee, J.H., et al., *Analyses of the National Institute on Aging Late-Onset Alzheimer's Disease Family Study: implication of additional loci.* Arch Neurol, 2008. **65**(11): p. 1518-26.
341. Lee, J.H., et al., *Expanded genomewide scan implicates a novel locus at 3q28 among Caribbean hispanics with familial Alzheimer disease.* Arch Neurol, 2006. **63**(11): p. 1591-8.
342. Lee, J.H., et al., *Fine mapping of 10q and 18q for familial Alzheimer's disease in Caribbean Hispanics.* Mol Psychiatry, 2004. **9**(11): p. 1042-51.
343. Myers, A., et al., *Full genome screen for Alzheimer disease: stage II analysis.* Am J Med Genet, 2002. **114**(2): p. 235-44.
344. Pericak-Vance, M.A., et al., *Complete genomic screen in late-onset familial Alzheimer disease. Evidence for a new locus on chromosome 12.* JAMA, 1997. **278**(15): p. 1237-41.
345. Sillen, A., et al., *Expanded high-resolution genetic study of 109 Swedish families with Alzheimer's disease.* Eur J Hum Genet, 2008. **16**(2): p. 202-8.
346. Sillen, A., et al., *Genome scan on Swedish Alzheimer's disease families.* Mol Psychiatry, 2006. **11**(2): p. 182-6.
347. Mayeux, R., et al., *Plasma A[beta]40 and A[beta]42 and Alzheimer's disease: relation to age, mortality, and risk.* Neurology, 2003. **61**(9): p. 1185-90.
348. Bagrodia, S., et al., *A novel regulator of p21-activated kinases.* J Biol Chem, 1998. **273**(37): p. 23633-6.
349. Feng, Q., et al., *Regulation of the Cool/Pix proteins: key binding partners of the Cdc42/Rac targets, the p21-activated kinases.* J Biol Chem, 2002. **277**(7): p. 5644-50.
350. Boesgaard, T.W., et al., *Variants at DGKB/TMEM195, ADRA2A, GLIS3 and C2CD4B loci are associated with reduced glucose-stimulated beta cell function in middle-aged Danish people.* Diabetologia, 2010.
351. Han, W. and C. Li, *Linking type 2 diabetes and Alzheimer's disease.* Proc Natl Acad Sci U S A, 2010. **107**(15): p. 6557-8.
352. Ueki, N., et al., *NOLP: identification of a novel human nucleolar protein and determination of sequence requirements for its nucleolar localization.* Biochem Biophys Res Commun, 1998. **252**(1): p. 97-102.
353. Wang, S.S., et al., *Identification of novel methylation markers in cervical cancer using restriction landmark genomic scanning.* Cancer Res, 2008. **68**(7): p. 2489-97.

354. Chen, W.M. and G.R. Abecasis, *Family-based association tests for genomewide association scans.* Am J Hum Genet, 2007. **81**(5): p. 913-26.
355. Katoh, M., *CLDN23 gene, frequently down-regulated in intestinal-type gastric cancer, is a novel member of CLAUDIN gene family.* Int J Mol Med, 2003. **11**(6): p. 683-9.
356. Lal-Nag, M. and P.J. Morin, *The claudins.* Genome Biol, 2009. **10**(8): p. 235.
357. Ouban, A. and A.A. Ahmed, *Claudins in human cancer: a review.* Histol Histopathol, 2010. **25**(1): p. 83-90.
358. Navarro, B., et al., *Nonselective and G betagamma-insensitive weaver K+ channels.* Science, 1996. **272**(5270): p. 1950-3.
359. Chung, H.J., et al., *G protein-activated inwardly rectifying potassium channels mediate depotentiation of long-term potentiation.* Proc Natl Acad Sci U S A, 2009. **106**(2): p. 635-40.
360. Fernandez-Medarde, A., et al., *Laser microdissection and microarray analysis of the hippocampus of Ras-GRF1 knockout mice reveals gene expression changes affecting signal transduction pathways related to memory and learning.* Neuroscience, 2007. **146**(1): p. 272-85.
361. Hamilton, G., et al., *Candidate gene association study of insulin signaling genes and Alzheimer's disease: evidence for SOS2, PCK1, and PPARgamma as susceptibility loci.* Am J Med Genet B Neuropsychiatr Genet, 2007. **144B**(4): p. 508-16.
362. Kimura, R., et al., *The DYRK1A gene, encoded in chromosome 21 Down syndrome critical region, bridges between beta-amyloid production and tau phosphorylation in Alzheimer disease.* Hum Mol Genet, 2007. **16**(1): p. 15-23.
363. Bullard, J.M., et al., *Expression and characterization of a human mitochondrial phenylalanyl-tRNA synthetase.* J Mol Biol, 1999. **288**(4): p. 567-77.
364. Adam, P.J., et al., *Comprehensive proteomic analysis of breast cancer cell membranes reveals unique proteins with potential roles in clinical cancer.* J Biol Chem, 2003. **278**(8): p. 6482-9.
365. Ghoussaini, M., et al., *Multiple loci with different cancer specificities within the 8q24 gene desert.* J Natl Cancer Inst, 2008. **100**(13): p. 962-6.
366. Martin, E.R., et al., *Interaction between the alpha-T catenin gene (VR22) and APOE in Alzheimer's disease.* J Med Genet, 2005. **42**(10): p. 787-92.
367. Chiba, H., et al., *Cloning of a gene for a novel epithelium-specific cytosolic phospholipase A2, cPLA2delta, induced in psoriatic skin.* J Biol Chem, 2004. **279**(13): p. 12890-7.
368. Tao, R., et al., *A family based study of the genetic association between the PLA2G4D gene and schizophrenia.* Prostaglandins Leukot Essent Fatty Acids, 2005. **73**(6): p. 419-22.
369. Rivadeneira, F., et al., *Twenty bone-mineral-density loci identified by large-scale meta-analysis of genome-wide association studies.* Nat Genet, 2009. **41**(11): p. 1199-206.

370. Styrkarsdottir, U., et al., *Multiple genetic loci for bone mineral density and fractures.* N Engl J Med, 2008. **358**(22): p. 2355-65.
371. Cantagrel, V., et al., *Disruption of a new X linked gene highly expressed in brain in a family with two mentally retarded males.* J Med Genet, 2004. **41**(10): p. 736-42.
372. Creutz, C.E., et al., *The copines, a novel class of C2 domain-containing, calcium-dependent, phospholipid-binding proteins conserved from Paramecium to humans.* J Biol Chem, 1998. **273**(3): p. 1393-402.
373. Maitra, R., et al., *Cloning, molecular characterization, and expression analysis of Copine 8.* Biochem Biophys Res Commun, 2003. **303**(3): p. 842-7.
374. Kajiho, H., et al., *RIN3: a novel Rab5 GEF interacting with amphiphysin II involved in the early endocytic pathway.* J Cell Sci, 2003. **116**(Pt 20): p. 4159-68.
375. Zou, F., et al., *Gene expression levels as endophenotypes in genome-wide association studies of Alzheimer disease.* Neurology, 2010. **74**(6): p. 480-6.
376. Prince, J.A., et al., *Genetic variation in a haplotype block spanning IDE influences Alzheimer disease.* Hum Mutat, 2003. **22**(5): p. 363-71.
377. Miners, J.S., et al., *Abeta-degrading enzymes in Alzheimer's disease.* Brain Pathol, 2008. **18**(2): p. 240-52.
378. Bertram, L., *Alzheimer's disease genetics current status and future perspectives.* Int Rev Neurobiol, 2009. **84**: p. 167-84.
379. Bertram, L., et al., *Evidence for genetic linkage of Alzheimer's disease to chromosome 10q.* Science, 2000. **290**(5500): p. 2302-3.
380. Schaid, D.J., et al., *Score tests for association between traits and haplotypes when linkage phase is ambiguous.* Am J Hum Genet, 2002. **70**(2): p. 425-34.
381. Purcell, S., et al., *PLINK: a tool set for whole-genome association and population-based linkage analyses.* Am J Hum Genet, 2007. **81**(3): p. 559-75.
382. Harman, D., *The aging process.* Proc Natl Acad Sci U S A, 1981. **78**(11): p. 7124-8.
383. Elstner, M., et al., *The mitochondrial proteome database: MitoP2.* Methods Enzymol, 2009. **457**: p. 3-20.
384. Wright, A.F., et al., *Lifespan and mitochondrial control of neurodegeneration.* Nat Genet, 2004. **36**(11): p. 1153-8.
385. Cantuti-Castelvetri, I., et al., *Somatic mitochondrial DNA mutations in single neurons and glia.* Neurobiol Aging, 2005. **26**(10): p. 1343-55.
386. Mambo, E., et al., *Electrophile and oxidant damage of mitochondrial DNA leading to rapid evolution of homoplasmic mutations.* Proc Natl Acad Sci U S A, 2003. **100**(4): p. 1838-43.
387. Simon, D.K., et al., *Somatic mitochondrial DNA mutations in cortex and substantia nigra in aging and Parkinson's disease.* Neurobiol Aging, 2004. **25**(1): p. 71-81.
388. King, M.P. and G. Attardi, *Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation.* Science, 1989. **246**(4929): p. 500-3.

389. Helbock, H.J., et al., *DNA oxidation matters: the HPLC-electrochemical detection assay of 8-oxo-deoxyguanosine and 8-oxo-guanine.* Proc Natl Acad Sci U S A, 1998. **95**(1): p. 288-93.
390. Cottrell, D.A., et al., *Mitochondrial enzyme-deficient hippocampal neurons and choroidal cells in AD.* Neurology, 2001. **57**(2): p. 260-4.
391. Parker, W.D., Jr., et al., *Electron transport chain defects in Alzheimer's disease brain.* Neurology, 1994. **44**(6): p. 1090-6.
392. Chandrasekaran, K., et al., *Decreased expression of nuclear and mitochondrial DNA-encoded genes of oxidative phosphorylation in association neocortex in Alzheimer disease.* Brain Res Mol Brain Res, 1997. **44**(1): p. 99-104.
393. Lundberg, K.S., et al., *High-fidelity amplification using a thermostable DNA polymerase isolated from Pyrococcus furiosus.* Gene, 1991. **108**(1): p. 1-6.
394. Tindall, K.R. and T.A. Kunkel, *Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase.* Biochemistry, 1988. **27**(16): p. 6008-13.
395. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
396. Wimo A & Prince, M., *The Global Impact of Dementia.* World Alzheimers Report 2010, 2010.
397. *2008 Alzheimer's disease facts and figures.* Alzheimers Dement, 2008. **4**(2): p. 110-33.
398. Brookmeyer, R., S. Gray, and C. Kawas, *Projections of Alzheimer's disease in the United States and the public health impact of delaying disease onset.* Am J Public Health, 1998. **88**(9): p. 1337-42.
399. Bartus, R.T., et al., *The cholinergic hypothesis of geriatric memory dysfunction.* Science, 1982. **217**(4558): p. 408-14.
400. Berger, W., et al., *Memantine inhibits [3H]MK-801 binding to human hippocampal NMDA receptors.* Neuroreport, 1994. **5**(10): p. 1237-40.
401. Marx, J., *Alzheimer's congress. Drug shows promise for advanced disease.* Science, 2000. **289**(5478): p. 375-7.
402. Knapp, M.J., et al., *A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. The Tacrine Study Group.* JAMA, 1994. **271**(13): p. 985-91.
403. Raskind, M.A., et al., *Galantamine in AD: A 6-month randomized, placebo-controlled trial with a 6-month extension. The Galantamine USA-1 Study Group.* Neurology, 2000. **54**(12): p. 2261-8.
404. Rogers, S.L., et al., *A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Donepezil Study Group.* Neurology, 1998. **50**(1): p. 136-45.
405. Rosler, M., et al., *Efficacy and safety of rivastigmine in patients with Alzheimer's disease: international randomised controlled trial.* BMJ, 1999. **318**(7184): p. 633-8.
406. Klaver, D.W., et al., *Is BACE1 a suitable therapeutic target for the treatment of Alzheimer's disease? Current strategies and future directions.* Biol Chem, 2010. **391**(8): p. 849-59.

407. Bergmans, B.A. and B. De Strooper, *gamma-secretases: from cell biology to therapeutic strategies.* Lancet Neurol, 2010. **9**(2): p. 215-26.
408. Das, P., et al., *Amyloid-beta immunization effectively reduces amyloid deposition in FcRgamma-/- knock-out mice.* J Neurosci, 2003. **23**(24): p. 8532-8.
409. Lemere, C.A. and E. Masliah, *Can Alzheimer disease be prevented by amyloid-beta immunotherapy?* Nat Rev Neurol, 2010. **6**(2): p. 108-19.
410. Lemere, C.A., et al., *Novel Abeta immunogens: is shorter better?* Curr Alzheimer Res, 2007. **4**(4): p. 427-36.
411. Bacanu, S.A., B. Devlin, and K. Roeder, *Association studies for quantitative traits in structured populations.* Genet Epidemiol, 2002. **22**(1): p. 78-93.
412. Bourgain, C. and E. Genin, *Complex trait mapping in isolated populations: Are specific statistical methods required?* Eur J Hum Genet, 2005. **13**(6): p. 698-706.
413. Devlin, B. and K. Roeder, *Genomic control for association studies.* Biometrics, 1999. **55**(4): p. 997-1004.
414. Dyer, T.D., et al., *The effect of pedigree complexity on quantitative trait linkage analysis.* Genet Epidemiol, 2001. **21 Suppl 1**: p. S236-43.
415. Scheschonka, A., Z. Tang, and H. Betz, *Sumoylation in neurons: nuclear and synaptic roles?* Trends Neurosci, 2007. **30**(3): p. 85-91.
416. Li, Y., et al., *Positive and negative regulation of APP amyloidogenesis by sumoylation.* Proc Natl Acad Sci U S A, 2003. **100**(1): p. 259-64.
417. Zhang, Y.Q. and K.D. Sarge, *Sumoylation of amyloid precursor protein negatively regulates Abeta aggregate levels.* Biochem Biophys Res Commun, 2008. **374**(4): p. 673-8.
418. Phillips, H.A., et al., *CHRNB2 is the second acetylcholine receptor subunit associated with autosomal dominant nocturnal frontal lobe epilepsy.* Am J Hum Genet, 2001. **68**(1): p. 225-31.
419. Conti, D.V., et al., *Nicotinic acetylcholine receptor beta2 subunit gene implicated in a systems-based candidate gene study of smoking cessation.* Hum Mol Genet, 2008. **17**(18): p. 2834-48.
420. De Fusco, M., et al., *The nicotinic receptor beta 2 subunit is mutant in nocturnal frontal lobe epilepsy.* Nat Genet, 2000. **26**(3): p. 275-6.
421. Cook, L.J., et al., *Candidate gene association studies of the alpha 4 (CHRNA4) and beta 2 (CHRNB2) neuronal nicotinic acetylcholine receptor subunit genes in Alzheimer's disease.* Neurosci Lett, 2004. **358**(2): p. 142-6.
422. Feulner, T.M., et al., *Examination of the current top candidate genes for AD in a genome-wide association study.* Mol Psychiatry, 2009.
423. Kawamata, J. and S. Shimohama, *Association of novel and established polymorphisms in neuronal nicotinic acetylcholine receptors with sporadic Alzheimer's disease.* J Alzheimers Dis, 2002. **4**(2): p. 71-6.
424. Law, S.F., et al., *The docking protein HEF1 is an apoptotic mediator at focal adhesion sites.* Mol Cell Biol, 2000. **20**(14): p. 5184-95.
425. Aquino, J.B., et al., *Differential expression and dynamic changes of murine NEDD9 in progenitor cells of diverse tissues.* Gene Expr Patterns, 2008. **8**(4): p. 217-26.

426. Vogel, T., et al., *Transforming growth factor beta promotes neuronal cell fate of mouse cortical and hippocampal progenitors in vitro and in vivo: identification of Nedd9 as an essential signaling component.* Cereb Cortex, 2010. **20**(3): p. 661-71.
427. Mehta, P.D., et al., *Plasma and cerebrospinal fluid levels of amyloid beta proteins 1-40 and 1-42 in Alzheimer disease.* Arch Neurol, 2000. **57**(1): p. 100-5.
428. Lettre, G., *Genetic regulation of adult stature.* Curr Opin Pediatr, 2009. **21**(4): p. 515-22.
429. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.
430. McCarthy, M.I., et al., *Genome-wide association studies for complex traits: consensus, uncertainty and challenges.* Nat Rev Genet, 2008. **9**(5): p. 356-69.
431. Repapi, E., et al., *Genome-wide association study identifies five loci associated with lung function.* Nat Genet, 2010. **42**(1): p. 36-44.
432. Shokolenko, I., et al., *Oxidative stress induces degradation of mitochondrial DNA.* Nucleic Acids Res, 2009. **37**(8): p. 2539-48.
433. Claesson, M.J., et al., *Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions.* Nucleic Acids Res, 2010. **38**(22): p. e200.
434. Coon, K.D., et al., *Quantitation of heteroplasmy of mtDNA sequence variants identified in a population of AD patients and controls by array-based resequencing.* Mitochondrion, 2006. **6**(4): p. 194-210.
435. Govan, H.L., 3rd, Y. Valles-Ayoub, and J. Braun, *Fine-mapping of DNA damage and repair in specific genomic segments.* Nucleic Acids Res, 1990. **18**(13): p. 3823-30.
436. Yakes, F.M. and B. Van Houten, *Mitochondrial DNA damage is more extensive and persists longer than nuclear DNA damage in human cells following oxidative stress.* Proc Natl Acad Sci U S A, 1997. **94**(2): p. 514-9.
437. Hunter, S.E., et al., *The QPCR assay for analysis of mitochondrial DNA damage, repair, and relative copy number.* Methods, 2010. **51**(4): p. 444-51.
438. Gudnason, H., et al., *Comparison of multiple DNA dyes for real-time PCR: effects of dye concentration and sequence composition on DNA amplification and melting temperature.* Nucleic Acids Res, 2007. **35**(19): p. e127.
439. Parker, W.D., Jr., C.M. Filley, and J.K. Parks, *Cytochrome oxidase deficiency in Alzheimer's disease.* Neurology, 1990. **40**(8): p. 1302-3.
440. Parker, W.D., Jr. and J.K. Parks, *Cytochrome c oxidase in Alzheimer's disease brain: purification and characterization.* Neurology, 1995. **45**(3 Pt 1): p. 482-6.
441. Van Goethem, G., et al., *Mutation of POLG is associated with progressive external ophthalmoplegia characterized by mtDNA deletions.* Nat Genet, 2001. **28**(3): p. 211-2.

442. Longley, M.J., et al., *Mutant POLG2 disrupts DNA polymerase gamma subunits and causes progressive external ophthalmoplegia.* Am J Hum Genet, 2006. **78**(6): p. 1026-34.
443. Spelbrink, J.N., et al., *Human mitochondrial DNA deletions associated with mutations in the gene encoding Twinkle, a phage T7 gene 4-like protein localized in mitochondria.* Nat Genet, 2001. **28**(3): p. 223-31.
444. Copeland, W.C., *Inherited mitochondrial diseases of DNA replication.* Annu Rev Med, 2008. **59**: p. 131-46.

Locus	Oligo type	Purpose	Name	Sequence
ND1	Primer	Cloning	ND1-F	CACCCAAGAACAGGGTTTGT
	Primer		ND1-R	TTTTGGATTCTCAGGATGG
HVR	Primer	Cloning	HVR-F	CACCATTAGCACCCAAAGCT
	Primer		HVR-R	CTGTTAAAAGTGCATACCGCC
D-loop	Primer	Cloning	D-loop-F	ACAGTCAAATCCCTTCTCGTC
	Primer		D-Loop-R	TGCGTGCTTGATGCTTGTT
COII	Primer	Cloning	COII-F	TCCATTTCACTATCATATTCATCG
	Primer		COII-R	GAGGTAGGTGGTAGTTTGTGTTTAA
16S	Primer	Cloning	16S-F	CAGACAACCTTAGCCAAACCA
	Primer		16S-R	TGGCCATGGGTATGTTGT
16s	Probe	qPCR	16s-Probe	CTGTCTCTTACTTTTAACCAGTG
	Primer		16s-For	TGGCTCCACGAGGGTTCA
	Primer		16s-Rev	TCTTCACGGGCAGGTCAATT
D-loop	Probe	qPCR	D-loop-Probe	CACCCCCCAACTAACACATTATTTTCCCC
	Primer		D-loop-For	TATCTTTTGGCGGTATGCACTTTTAACAGT
	Primer		D-loop-Rev	TGATGAGATTAGTAGTATGGGAGTGG
COII	Probe	qPCR	COII-Probe	CAATTCCCGGACGTCTAAACCAAACCACTTTC
	Primer		COII-For	CCCCACATTAGGCTTAAAAACAGAT
	Primer		COII-Rev	TATACCCCCGGTCGTGTAGCGGT

Table i. Primers and probes used in the studies described in chapter 4.